

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification<sup>6</sup> : <b>C12N 15/12, 15/55, 15/80, 15/85, 9/16, 1/15, C07K 14/37, 14/47, 1/22, 16/14, 16/18, G01N 33/50, A61K 38/16, 38/17</b></p>	<p><b>A1</b></p>	<p>(11) International Publication Number: <b>WO 98/36064</b></p> <p>(43) International Publication Date: 20 August 1998 (20.08.98)</p>
<p>(21) International Application Number: PCT/US98/02709</p> <p>(22) International Filing Date: 13 February 1998 (13.02.98)</p> <p>(30) Priority Data: 60/039,738 14 February 1997 (14.02.97) US</p> <p>(71) Applicants (for all designated States except US): THE PROCTER &amp; GAMBLE COMPANY [US/US]; One Procter &amp; Gamble Plaza, Cincinnati, OH 45202 (US). UNIVERSITY OF SHEFFIELD MED. SCHOOL [-/GB]; Beech Hill Road, Sheffield S10 2RX (GB).</p> <p>(72) Inventors; and</p> <p>(75) Inventors/Applicants (for US only): COOK, Jonathan, Shaun [GB/US]; 2984 Pine Grove Lane, Maineville, OH 45039 (US). EBETINO, Frank, Hallock [US/US]; 11249 Acrewood Drive, Cincinnati, OH 45249 (US). IBBOTSON, Kenneth, John [CN/US]; 8899 Appleseed Drive, Cincinnati, OH 45249 (US). JI, Xiaohui [CN/US]; 8495F Jonathan Lane, Maineville, OH 45039 (US). ROGERS, Michael, John [GB/GB]; 19 Old Castle, Slains, By Collieston, Aberdeenshire AB41 8ST (GB). WATTS, Donald, Jeremy [GB/GB]; The Bungalow, 29 Hallam Gate Road, Sheffield S10 2TN (GB). RUSSELL, Roger, Graham, Goodwin, Rus-</p>	<p>sell [-/GB]; Sheffield University Medical School, Beech Hill Road, Sheffield S10 2RX (GB). XIONG, Xiao, Juan [CN/CA]; 1066 Stormont Street, Ottawa, Ontario K2C 0M9 (CA).</p> <p>(74) Agents: REED, T., David et al.; The Procter &amp; Gamble Company, 5299 Spring Grove Avenue, Cincinnati, OH 45217 (US).</p> <p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b></p> <p><i>With international search report.</i></p> <p><i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>	
<p>(54) Title: PHOSPHONATE BINDING PROTEINS</p> <p>(57) Abstract</p> <p>The present invention relates to proteins capable of binding bisphosphonate or bisphosphonate analogues, to methods for producing and identifying these proteins, to their corresponding genes and to various uses of the proteins, for example, in therapy and in the screening, isolation, synthesis, design and evaluation of bisphosphonate-based drugs.</p>		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## PHOSPHONATE BINDING PROTEINS

### TECHNICAL FIELD

The present invention relates to proteins capable of binding bisphosphonate or bisphosphonate analogues, to methods for producing and identifying these proteins, to their corresponding genes and to various uses of the proteins, for example, in therapy and in the screening, isolation, synthesis, design and evaluation of bisphosphonate-based drugs.

### BACKGROUND

#### Bone pathology

A number of diseases are recognized which arise from bone destruction or disorders of bone metabolism. These diseases are of great clinical importance and have been the subject of intense scientific research for several decades.

Bone destruction can result from various cancers and from rheumatoid arthritis. Metabolic bone disorders commonly involve excessive bone resorption and include Paget's disease, hypercalcaemia (both tumour-induced and non-tumour induced), bone metastases and osteoporosis.

Paget's disease (a focal increase in bone turnover) is fairly common, and in some countries affects up to 5% of the population over 50 years of age. The disease may be caused by a slow virus infection, and leads to bone pain, deformities and fractures.

Bone metastases can induce bone destruction either through local invasion or via the secretion of bone-resorbing agents into the blood stream.

Hypercalcaemia can result either from an increase in the flow of calcium from bone or the intestine to the blood, or from an increase in tubular reabsorption of calcium in the kidney. It can induce a wide variety of physiological and/or pharmacological disturbances and can be life threatening.

Osteoporosis is characterized by a reduction in the quantity of bone. This loss of bone tissue often causes mechanical failure, and bone fractures frequently occur in the hip and spine of women suffering from postmenopausal osteoporosis. Kyphosis (abnormally increased curvature of the thoracic spine) is another common feature.

Two types of osteoporosis are recognized: primary and secondary. Secondary osteoporosis is the result of an identifiable disease process or agent, while primary osteoporosis (which constitutes about 90% of all cases) includes postmenopausal osteoporosis, age-associated osteoporosis (affecting a majority of individuals over the age of 70) and idiopathic osteoporosis affecting middle-aged and younger men and women.

The mechanism of bone loss in osteoporosis is believed to involve an imbalance of the process of "bone remodeling". Bone remodeling occurs throughout life, renewing the skeleton and maintaining the strength of bone. This remodeling involves the erosion and filling of discrete sites on the surface of bones, by an organized group of cells called "basic multicellular units" or "BMUs". BMUs primarily consist of osteoclasts, osteoblasts, and their cellular precursors. In the remodeling cycle, bone is resorbed at the site of an "activated" BMU by an osteoclast, forming a resorption cavity. This cavity is then filled with bone by osteoblasts.

Normally, in adults, the remodeling cycle results in a small deficit in bone, due to incomplete filling of the bone resorption cavity. Thus, even in healthy adults, age-related bone loss occurs. However, in many people, particularly in postmenopausal osteoporotics, there is an increase in the number of BMUs that are activated. This increased activation accelerates bone remodeling, resulting in abnormally high bone loss.

Many compositions and methods are described in the medical literature for the treatment of the above-described diseases and conditions, and most attempt to either slow the loss of bone or produce a net gain in bone mass.

Administration of oestrogen has been used as a means both to prevent and to treat osteoporosis in postmenopausal women. However, the use of oestrogen has been associated with certain side effects, such as uterine bleeding.

Other treatments are based on the administration of parathyroid hormone.

The hormone calcitonin has also been used to treat Paget's disease (and to a lesser extent tumour bone disease), and can be effective in decreasing bone turnover.

However, the incidence of relapse is high, and side effects limit the therapeutic usefulness of calcitonin.

Perhaps one of the most successful class of drug for the treatment of the above diseases has proved to be the bisphosphonates.

#### Bisphosphonates

Inorganic pyrophosphate has high affinity for bone mineral and is able to inhibit the precipitation and dissolution of calcium phosphate crystals in vitro and to inhibit bone mineralization in vivo. These activities are thought to arise from direct physicochemical effects (such as adsorption to hydroxyapatite, inhibition of dissolution of hydroxyapatite and crystal growth inhibition). Pyrophosphate has therefore found application as an antitartar agent for use in toothpastes.

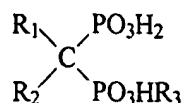
However, inorganic pyrophosphate (PPi) is rapidly hydrolysed following administration (particularly oral administration) due to the presence of the relatively labile phosphorus-oxygen bond P-O-P. This severely limits their pharmaceutical utility, and has prompted a search for PPi analogues which exhibit similar physicochemical activities while resisting enzymatic hydrolysis in vivo.

Bisphosphonates, which are characterized by phosphorus-carbon (P-C-P) bonds, are stable analogues of naturally occurring inorganic pyrophosphates which to a great extent overcome the limitations associated with inorganic pyrophosphates. Bisphosphonates are resistant to chemical and enzymatic hydrolysis but retain the therapeutic activity of PPi.

Unlike pyrophosphates, however, bisphosphonates exhibit properties which extend beyond those attributable to purely physicochemical phenomena. In particular, bisphosphonates have been found to be inhibitors of osteoclast-mediated bone resorption in organ cultures of bone and in animal models. Bisphosphonates therefore have broader clinical utility than PPi, and have found widespread application in several main clinical areas, e.g., (a) as bone imaging agents for diagnostic purposes, usually in the form of <sup>99m</sup>technetium derivatives, (b) as anti-resorptive agents to combat bone loss associated with Paget's disease, hypercalcaemia associated with malignancy and bone metastases and osteoporosis, (c) as calcification inhibitors in patients with ectopic calcification and ossification, and (d) as antitartar agents for use in toothpastes.

Bisphosphonates are now among the most important therapeutic agents for the treatment of pathological disorders of bone metabolism, including osteoporosis. Moreover, since some bisphosphonates appear to have anti-inflammatory as well as anti-resorptive effects in vivo, they may also have utility in the treatment of inflammation and rheumatoid arthritis.

Bisphosphonates described in the literature generally have the structure:



Without wishing to be bound by any theory, bisphosphonates appear to contain a "bone binding moiety" (the P-C-P group) and a "bioactive moiety", R<sub>2</sub>. The "bone binding moiety" appears to endow the bisphosphonate compound with the ability to adsorb to bone and/or to hydroxyapatite (a model for bone), while the R<sub>2</sub> bioactive moiety appears to determine the potency of the bisphosphonate.

Minor alterations to R<sub>2</sub> can have a marked effect on potency, and this property is therefore specific to R<sub>2</sub>. Varying the R<sub>2</sub> side chain has led to dramatic variations in potency (in some cases, 4 orders of magnitude) (Rogers et al., 1995, *Mol. Pharm.* 47:398-402).

R<sub>1</sub> is a moiety that assists in binding to bone or in bioactivity (as R<sub>2</sub>), R<sub>3</sub> is H or alkyl. Such compounds are described, for example, in U.S. Patent 5,391,743; Published European Patent application 186405 (published July 2, 1986); Published European Patent application 298553 (published January 1, 1989); published PCT applications WO 93/05044 (published December 9, 1993), WO 93/04469 (published December 9, 1993), WO 93/04979 (published December 9, 1993), WO 93/04978 (published December 9, 1993), and WO 93/04977 (published December 9, 1993), hereby incorporated by reference.

The literature also describes bisphosphonate analogs, sometimes also referred to as monophosphate analogues, where one phosphonate is replaced by a carboxylate, sulfonate or other acid or ester, such as in published applications WO 93/04993 (published December 9, 1993), and WO 93/04976 (published December 9, 1993), hereby incorporated by reference.

In, addition there may exist bisphosphonate "mimetics," which may have some, or none of the analogue's characteristics, yet bind to the bisphosphonate binding protein, and this may be useful in treating the same maladies as the bisphosphonates.

However, despite widespread recognition of the importance of bisphosphonates, the mechanism of action of these compounds has not been elucidated. It has been suggested that bisphosphonates may affect the differentiation and recruitment of osteoclast precursors or alter the capacity of mature osteoclasts to resorb bone by altering the permeability of the osteoclast membrane to small ions. Another hypothesis is that they act by affecting lysosomal enzyme production or cell metabolism or through toxic effects on osteoclasts (Carano et al., 1990, *J. Clin.*

Invest. 85:456-461). A further suggestion is that other cells in the bone microenvironment that regulate the activity of osteoclasts are involved in the antiresorptive mechanism.

There exists considerable opportunity for further innovation and development in this field. In addition, many further clinical applications may exist requiring different profiles of activity.

There is therefore a need to understand the mechanism of action of the bisphosphonates, to rapidly and efficiently screen potentially improved bisphosphonate drugs and to identify and create compounds having improved therapeutic activity.

The present inventors have now recognized that bisphosphonate drugs exert their effect (at least in part) via interaction with specific target proteins (hereinafter referred to as bisphosphonate binding proteins).

#### SUMMARY OF THE INVENTION

According to the present invention there is provided an isolated bisphosphonate binding protein(s), or homologues, fragments, muteins, equivalents or derivatives (e.g. fusion derivatives or synthetic peptides) thereof which substantially retain bisphosphonate binding activity.

The present invention also relates to bisphosphonate binding proteins that are useful in therapy and in the identification, isolation and/or design of novel drugs.

The invention also relates to isolated DNA encoding the bisphosphonate binding protein (or derivatives, fragments, etc.) thereof. The invention also relates to nucleic acid probes which are selectively hybridizable with the DNA of the invention.

The invention also relates to a method for producing a bisphosphonate binding protein.

The invention also relates to a method for designing and synthesizing a therapeutically active bisphosphonate or a mimetic thereof using a three-dimensional model of the bisphosphonate binding protein, or the bisphosphonate binding site of such protein

The invention also relates to the bisphosphonate binding protein, antibody thereto, mimetic or antagonist thereof for use in therapy, diagnosis or testing, both in vivo and in vitro.

The invention also relates to the protein of the invention for use in a method of screening bisphosphonates for therapeutic activity. Such screening contemplates test kits or screening kits, comprising the protein.

The invention also relates to the protein of the invention which is labeled for use in the test kits of the invention.

Also contemplated by the invention is a host cell comprising the vector of the invention.

Further aspects of the invention will become apparent as the description proceeds.

#### BRIEF DESCRIPTION OF THE DRAWINGS

The invention is described in further detail by way of DRAWINGS, listed below.

Figure 1: Affinity chromatography of a cell extract of Dictyostelium discoideum on an AHBuBP-affinity column.

Figure 2: Profile of eluted proteins from the AHBuBP-affinity column.

Figure 3 (SEQ. ID NO. 1 and SEQ. ID NO. 2): Nucleotide and deduced amino acid sequence of the cDNA for wild-type DP1. Since this cDNA is identical to the sequence of discoidin II, it is clear that DP1 is in fact discoidin II. The nucleotide sequence shown is that of the 968-bp cDNA of DP1 from wild-type D. discoideum Ax-2. The adenine base of the ATG initiation sequence is assigned as 1 in the numbering. Nucleotides are numbered in the right margin and amino acids on the left. An open reading frame of 771-bp encoding 257 amino acids is shown with a single-letter code for the translated amino acids. A termination codon (TAA) at the end of the coding sequence is marked with an asterisk. A putative signal peptide at the beginning of the amino acid sequence is indicated in italics. The nucleotide sequence of the PCR fragment XJ-450 used to screen the cDNA library is underlined. Multiple polyadenylation signal sequences (AATAAA) are shown in italics.

Figure 4: Comparison of the deduced amino acid sequences for DP1 (SEQ. ID NO. 3) and discoidin IA (DISC IA) (SEQ. ID NO. 4). Asterisks indicate positions of identity while dots indicate positions of conservation. Amino acid sequences obtained for peptides of DP1 are underlined and the corresponding numbers are shown in parentheses underneath the sequences. Regions used for generating primers XJ-1 (SEQ. ID NO. 5) and XJ-2 (SEQ. ID NO. 6) are in italics. The alignment is performed by use of the multiple alignment program of CLUSTALV (Higgins et al., 1992, *Comp Appl Biosci* 8: 189-191).

Figure 5: Northern blot analyses of DP1 mRNA expression in axenic strains. 0.5 ug samples of mRNA were fractionated on formaldehyde gels. After transfer to Hybond-N membrane (Amersham, Bucks, UK), the samples were hybridized to <sup>32</sup>P-labeled DNA probes. In (a), mRNAs on the blot were subjected to successive



hybridizations with 32P-labeled PCR fragments of DP1 (XJ-450) from sequence ID #1, discoidin IA sequence ID #6 and Dd-tcp1 (control). Lane 1 shows axenically-grown Ax-2, lane 2 shows bacterially-grown Ax-2. The columns on the right hand side represent the relative abundance of the hybridized mRNA transcripts in the corresponding lanes of the blots. In (b), hybridization to 32P-labeled DP1 cDNA XJ-450 of mRNA isolated from amoebae of strain Ax-2 harvested from axenic culture at low density ( $1 \times 10^5$  cells/ml) (A) and at high density ( $4 \times 10^6$  cells/ml)(B).

**Figure 6:** Southern blot analysis of DP1. Genomic DNA from the wild-type Ax-2 strain is subjected to restriction digestion and the products were separated on a 1% agarose gel (a). After transfer to Hybond-N membrane, the samples were hybridized to the 32P-labeled XJ-450 probe (b) or the discoidin IA probe (c). Note that in each lane there is only a single band of hybridization to the DP1 (discoidin II) cDNA XJ-450 whereas there were multiple bands for hybridization to the discoidin IA cDNA probe.

**Figure 7:** Sequence comparison matrices for DP1 (horizontal) with (a) human coagulation factor V, Wood et al., Nature 312:330-337 (1984); Jenny, et al., Proc. Natl. Acad. Sci., USA 84:4846-4850 (1987), (b) human coagulation factor VIII, Wood et al., Nature 312:330-337 (1984), (c) ORF7 linked to the *Rhodopseudomonas blastica* *atp* operon (Yat7-Rhob1) Tybulewicz et al., J. Mol. Biol. 179:185-214 (1984) and (d) milk fat globule (MFG) protein Larocca et al. 1991. The matrices were plotted using MDM78 mutation data matrix Pam250 as the scoring system (Schwartz and Dayhoff, 1978, Atlas of Protein Sequence and Structure. 5: 353-358, National Biomedical Research Foundation, Washington DC.), with a window size of 8 and a minimum score of 50%.

**Figure 8:** Hydrophilic plot and secondary structure prediction for DP1. The profile is constructed using the Kyte-Doolittle algorithm (Kyte and Doolittle, 1982. J. Mol. Biol. 157, 105-132) with a sliding window of seven amino acids. Values above the zero axis correspond to hydrophilic segments. Secondary structure predictions based on algorithms of both Chou-Fasman (Adv. Enzymol. Relat. Areas Mol. Bio., 47, 45-148, 1978) (CH, shadowed boxes) and Robson-Garnier (Garnier et al., 1978, J. Mol. Biol. 120, 97-120) (RG, filled boxes) are shown in the lower half of the Figure.

**Figure 9:** Prediction of a probable cleavage site for a signal peptide of DP1 by the method of van Heijne (1986) using the programme MacProt (Markiewicz, 1991, *BioTechniques* 10: 756-757+760+762-763; Luttko, 1990, *Comp. Method. Prog. Biomed.* 31:105-112). A window size of 15 residues of weight matrix for eukaryotic proteins is used. Generally, a protein having a signal peptide has one segment scoring greater than + 3.5 while cytosolic proteins have a score less than + 3.5.

**Figure 10:** Surface probability of the RGD-containing region of DP1. A surface probability profile of residues 76-88 is constructed using the Janin *et al.* (*J. Mol. Biol.* 125: 357-386, 1978) and Emini *et al.* (*J. Virol.* 55: 836-839, 1985) algorithms. A probability value above 0.5 is assigned to a water-accessible, "exposed" sequence and values below 0.5 to "buried" segments.

**Figure 11.** N-terminal amino acid sequence alignment of hDP1 (SEQ. ID NO. 7) and the rat round spermatid 29,000 Mr protein (RSP-29) (SEQ. ID NO. 8) Onoda and Djaliw 1993A. Asterisks indicate positions of identity, colons indicate positions of conservation and dots positions where the two sequences contain different amino acids.

**Figure 12:** Schematic representation of the amplification of hDP1 cDNA by the polymerase chain reaction. Degenerate primers DJ1, DJ5, DJ9 (sense) and DJ10 (antisense) were designed from a knowledge of the peptide sequences of hDP1. The nucleotide sequence of the primer UAP (antisense) had been incorporated into the first strand cDNA during reverse transcription using primer oligo(dT)17-AP.

**Figure 13:** Nucleotide sequence and deduced amino acid sequence of hDP1 cDNA. The nucleotide sequence of hDP1 (SEQ. ID NO. 9 and SEQ. ID NO 10) cDNA isolated from a human testis cDNA library consists of 1161 base pairs. The largest open reading frame consists of 936 bp and contains six ATG codons. The deduced amino acid sequence is shown in a single letter code. The mature hDP1 protein appears to be encoded by the nucleotide sequence started from the second ATG codon and contains 260 amino acids (shown in bold type). The adenine base of this second ATG codon and the methionine encoded by this codon are assigned as 1 respectively in the numbering. Nucleotides are numbered in the right margin and amino acids in the left. A putative N-terminal sequence of 48 amino acids starting from the first ATG codon is shown in italics. A stop codon (TGA) at the terminus of the translation sequence is marked with an asterisk. The amino acid sequences of peptides 1 and 2 are underlined.

**Figure 14:** Northern blot analyses of hDP1 in human tissues. Human multiple

tissue blots (I and II) containing 2 ug poly(A<sup>+</sup>)RNA from various tissues were purchased from Clontech (Palo Alto, California). Northern blots III and IV were prepared by running 1 ug of poly(A<sup>+</sup>)RNA in a formaldehyde gel and blotting onto Hybond-N membrane. Hybridization is carried out using a [<sup>32</sup>P]-labeled PCR fragment of hDP1 cDNA. A DNA fragment of 3-actin is also labeled with [<sup>32</sup>P] and hybridized to the same blots. The height of the bar on top of each lane represents the relative abundance of hDP1 mRNA in that tissue.

**Figure 15:** Southern analysis of hDP1 on a zoo-blot. The Southern blot containing 8 ug of genomic DNA per lane from nine eukaryotic species is purchased from Clontech. The DNA had been digested with EcoRI, run on a 0.7% agarose gel and transferred to a nylon membrane. Hybridization is carried out using a [<sup>32</sup>P]-labeled PCR fragment of hDP1 (SEQ. ID NO. 13) cDNA. Lanes 1-9 contain, in order, genomic DNA from yeast, chicken, rabbit, cow, dog, mouse, rat, monkey and man.

**Figure 16:** Northern blot analyses of hDP1 homologues in Dictyostelium discoideum. A [<sup>32</sup>P]-labeled PCR fragment of hDP1 cDNA is hybridized to a northern blot containing 5 ug poly(A<sup>+</sup>)RNA isolated from Dictyostelium amoebae: Lane 1, strain Ax-2 grown with bacteria; Lane 2 strain Ax-2 grown axenically.

**Figure 17:** Sequence alignment of hDP1. (a) and (b) show matrix plots of hDP1 (horizontal) against ORF3 linked to the Rhodospseudomonas blastica atp operon (ATPase-ORF3, vertical) and ORF1 linked to the genes for arginyl tRNA synthetase and ribonuclease H of Buchnera aphidicola (ATS-ORF1, vertical). The matrices were plotted using MDM78 mutation data matrix Pam250 as the scoring system (Schwartz and Dayhoff, 1978, *Atlas of Protein Sequence and Structure*, Vol. 5, 353-358, National Biomedical Research Foundation, Washington, DC), with a window size of 8 and a minimum score of 55%. In (c), the sequence alignment of hDP1 with ATPase-ORF3 and ATS-ORF1 is shown. The alignment is performed by using multiple alignment program CLUSTALV (Higgins et al., 1992, *Comp. Appl. Biosci.* 8: 189-191). A position where three sequences contain the same amino acid is indicated by an asterisk. A dot indicates a position where the three sequences contain amino acids having similar properties.

**Figure 18.** Sequence comparison of hDP1 (SEQ. ID NO. 10) with aspartate aminotransferase (AAT) (SEQ. ID NO. 11) from chicken mitochondria. hDP1 is shown in the upper line and DP1 in the lower one. "|" indicates identity between aligned residues; ":" indicates similarity. The comparison is performed using the program "bestfit" of the GCG package (Devereux et al., 1984, *Nucleic Acids Res.* 12: 387-395). Identity between the two proteins is 14.3% and the similarity is 44.1%.

**Figure 19:** Hydrophilic plot and secondary structure prediction for hDP1. The hydrophilic profile is constructed using the Kyte-Doolittle algorithm (Kyte and Doolittle, 1982, *J. Mol. Biol.* 157, 105-132) with a sliding window of seven amino acids. Values above the zero axis correspond to hydrophilic segments. Secondary structure prediction is based on algorithms of both Chou-Fasman (*Adv. Enzymol. Relat. Areas Mol. Biol.* 47, 45-148, 1978) (CH, shadowed boxes) and Robson-Garnier (Garnier et al., 1978, *J. Mol. Biol.* 120, 97-120) (RG, filled boxes), shown in the lower half of the Figure.

**Figure 20.** Amino acid sequence comparison of hDP1 (SEQ. ID NO. 10) with DP1 (SEQ. ID NO. 3) (discoidin II). hDP1 is shown in the upper line and DP1 in the lower. "|" indicates identity between aligned residues; ":" indicates similarity. The comparison is performed using the program "bestfit" of the GCG package. (Devereux et al., 1984, *Nucleic Acids Res.* 12: 387-395). The identity between the two proteins is 12.8% and the similarity is 38.5%.

**Figure 21.** Nucleotide and predicted amino acid sequence of DdCyP2 cDNA (SEQ. ID NO. 12). The nucleotide sequence shown is a 685 bp cDNA of DdCyP2 (SEQ. ID NO. 13) isolated from a *Dictyostelium discoideum* strain Ax2 cDNA library. Nucleotides are numbered in the left margin and amino acids on the right. An open reading frame of 540 bp encoding 180 amino acids is shown using a single letter code for the translated amino acids. A start codon (ATG) and stop codon (TAA) are underlined. An RGD motif is in bold.

**Figure 22.** Alignment of the amino acid sequence of Dd CyP2 and the sequences of selected members of the cyclophilin A family of proteins. The sequences were extracted from the updated releases from GenBank and Swissprot. The alignment is performed using "pileup" of the GCG program (the Wisconsin Genetic Computer Group). The sequences of the CyPs were derived from the following sources: (1) Dd CyP2; (2) Dd CyP1 (Barisic et al., 1991, *Developmental Genetics* 12:50-53); (3)

*Brassica napus* (p24525, Gasser et al., 1990, *Proc. Natl Acad. Sci. USA* 87: 9519-9532); (4) *Saccharomyces cerevisiae* (p14832, Haendler et al., 1989, *Gene* 83: 39-46); (5) human (p05092, Haendler et al., 1987, *EMBO J.* 6: 947-950). Residues of human CyPA that are located in close contact with a tetrapeptide Ala-Ala-Pro-Ala substrate are marked with an asterisk (Kallen & Walkinshaw, 1992, *FEBS Lett.* 300: 286-290); residues of human CyPA that are in close contact with bound cyclosporin A are marked with a cross (Theriault et al., 1993, *Nature* 361: 88-91; Pflugl et al., 1993, *Nature* 361: 91-94).

Figure 23. (SEQ. ID NO. 13) Alignment of the amino acid sequences of Dd CyP2 and four human cyclophilins. The sequences were extracted from the updated releases from GenBank and Swissprot. The alignment is performed using "pileup" of the GCG program (the Wisconsin Genetic Computer Group). The sequences of the CyPs were derived from the following sources: (1) human CyPA (p05092, Haendler et al., 1987, *EMBO J.* 6: 947-950), (2) human CyPD (p30405, Bergsma et al., 1991, *J. Biol. Chem.* 266: 23204-23214); (3) Dd CyP2; (4) human CyPB (p23284, Price et al., 1991, *Proc. Natl. Acad. Sci.* 88: 1903-1907); (5) human CyPC (Schneider et al., 1994, *Biochemistry* 33: 8218-8224).

Figure 24. (SEQ. ID NO. 13) Alignment of the amino acid sequences of Dd CyP2 and CyPBs. The sequences were extracted from the updated releases from GenBank and Swissprot. The alignment is performed using "pileup" of the GCG program (the Wisconsin Genetic Computer Group). The sequences of the CyPBs were derived from the following sources: (1) human (p23284, Price et al., 1991, *Proc. Natl. Acad. Sci.* 88: 1903-1907); (2) mouse (p24369, Hasel et al., 1991, *Mol. and Cellular Bio.* 11: 3484-3491); (3) chick (p24367, Caroni et al., 1991, *J. Biol. Chem.* 266: 10739-10742); (4) rat (p24368, Iwai and Inagami et al., 1990, *Kidney Int.* 37: 1460-1465). (5) Dd CyP2 (XP1). Potential hydrophobic signal sequences of CyPBs are underlined. The RGD motifs are in bold-type.

Figure 25. (SEQ. ID NO. 13) Alignment of the amino acid sequences of two Dd CyPs and some plant CyPs. The sequences were extracted from the updated releases from GenBank and Swissprot. The alignment is performed using "pileup" of the GCG program (the Wisconsin Genetic Computer Group). The sequences for the CyPs were derived from the following sources: (1) *Arabidopsis thaliana* (l14844, Lippuner et al., 1994, *J. Biol. Chem.* 269: 7863-7868) (2) tomato (m55019, Gasser et al., 1990, *Proc. Natl. Acad. Sci. USA* 87: 9510-9523); (3) *Brassica napus* (m55018, Gasser et al., 1990, *Proc. Natl. Acad. Sci. USA* 87: 9510-9523); (4) onion (l13365, Clark et al., 1993, direct submission of the onion cyclophilin to the GenBank); (5) maize (m55021, Gasser et al., 1990, *Proc. Natl. Acad. Sci. USA* 87:

9510-9523); (6) Dd CyP2 (XP1); (7) *Arabidopsis thaliana* (x63616, Bartling et al., 1991, *Plant Mol. Biol.* 19: 529-530); (8) *Arabidopsis thaliana* (114845, nuclear-encoded chloroplast stromal, Lippuner et al., 1994, *J. Biol. Chem.* 269: 7863-7868); (9) Dd CyP1 (Barisic et al., 1991, *Developmental Genetics* 12:50-53). The seven amino acid insertion in Dd CyP2 is in bold-type. The potential ATP/GTP binding sites are underlined.

### DETAILED DESCRIPTION

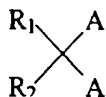
The term "isolated" is used herein to indicate that the binding protein is substantially isolated with respect to the complex cellular milieu in which it naturally occurs. The absolute level of purity is not critical, and those skilled in the art can readily determine appropriate levels of purity depending upon the intended use for the protein. In many circumstances, the isolated bisphosphonate binding protein will form part of a composition, buffer system or pharmaceutical excipient, which may for example contain other components (including other proteins, such as albumin). In other circumstances, the bisphosphonate binding protein may be purified to essential homogeneity, for example as determined by PAGE or column chromatography (for example HPLC).

The term "bisphosphonate binding protein" is used herein to define a protein which can bind bisphosphonate and/or act (either directly, or indirectly) as a target for bisphosphonate(s) in vivo. The bisphosphonate binding proteins of the invention may bind specifically to bisphosphonates, or may exhibit broader binding affinity and bind to other molecules in addition to bisphosphonate (with either the same or with a different affinity).

Thus, as defined herein, the bisphosphonate binding proteins of the invention may be targets of the physiological and/or pharmacological action of bisphosphonates and may mediate the physiological and/or pharmacological effects of bisphosphonates in vivo (such as growth inhibition in Dictyostelium discoideum and antiresorptive action in humans), either alone or as part of a complex comprising other proteins and/or molecules. Accordingly, the bisphosphonate binding proteins of the invention may be bisphosphonate receptors, and may be involved in signal transduction during a cellular response to bisphosphonate. The bisphosphonate binding protein may also be an enzyme for which bisphosphonates are substrate analogues.

The term "bisphosphonate" is used herein in a broad sense to cover not only bisphosphonates sensu stricto, as defined by the literature above, but also bisphosphonate analogues. Bisphosphonate analogues are those ligands which can

compete with osteoactive bisphosphonate for binding to the cellular targets of the osteoactive bisphosphonates, or which can compete *in vitro* (for example, on an affinity column) with osteoactive bisphosphonate (in either the free state or in the form of a derivative linked to an affinity column) for binding to the binding proteins of the invention. These include  $\text{HOOC-C-PO}_3\text{H}_2$ ,  $\text{HO}_3\text{S-C-PO}_3\text{H}_2$ ,  $\text{RHO}_3\text{P-C-PO}_3\text{H}_2$  and the like. Such compounds are disclosed for example in F. H. Ebetino, et al. U. S. Patent 4,868,164, issued July 6, 1988, F. H. Ebetino, et al. Published European Patent application 87/0274158, published July 13, 1988, F. H. Ebetino, et al. U. S. Patent 5,334,586, issued December 1, 1991, F. H. Ebetino, et al. U. S. Patent 4,868,164, issued August 2, 1994, C. N. Yu, et al. Published European Patent application 92/918467.9, published March 3, 1993, F.H. Ebetino, et al. Published PCT Patent application 93/04976, published December 9, 1993, all incorporated herein by reference. Bisphosphonates, and bisphosphonate analogues referred to throughout this application are:



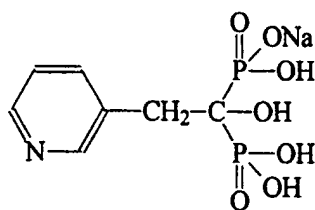
where

$\text{R}_2$  is a "bioactive moiety" as described above;

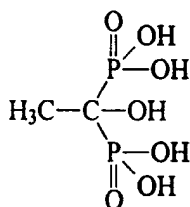
$\text{R}_1$  is a moiety that assists binding to bone, or assists bioactivity; and

each A is individually an acidic moiety, an ester or the like; or any moiety that binds to bone.

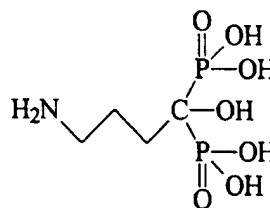
Specific examples of these include:



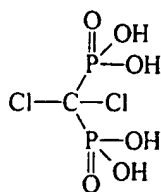
3PHEBP  
(Risedronate)



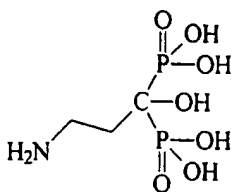
HEBP  
(Etidronate)



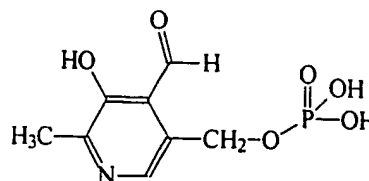
AHBuBP  
(Alendronate)



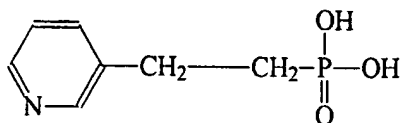
Cl<sub>2</sub>MBP  
clodronate



AHPBP  
(pamidronate)



PLP  
(pyridoxal phosphate)



NE10788

Bisphosphonate analogues include, but are not limited to pyridoxal phosphate (PLP), O-phosphorylethanolamine, O-phosphorylcholine, phosphatidyl ethanolamine and phospholipid bisphosphonate analogues. Pyridoxal phosphate (PLP), O-phosphorylethanolamine, O-phosphorylcholine, phosphatidyl ethanolamine are known in the art.

"Osteoactive bisphosphonates" are those bisphosphonates which act as antiresorptive drugs *in vivo*.

The term "Bisphosphonate mimetic" as defined herein, is a functional term that describes any molecule which performs similarly to a bisphosphonate *in vivo*, or *in vitro*. The structure of such molecules are secondary to their function.

In preferred embodiments of the invention, the bisphosphonate binding protein is a cyclophilin, a discoidin or a round spermatid protein (or homologue thereof).

The term "discoidin" is a term of the art, and defines a family of proteins sharing sequence similarity or homology with the discoidin proteins of *Dictyostelium* (particularly discoidin I) which are cytoplasmic components. The term discoidin covers discoidin homologues (e.g. mammalian homologues) which share conserved domains characteristic of the discoidin family.

The term "cyclophilin" (CyP) is also a term of the art and defines a family of proteins falling into at least two groups: the cyclophilin A group and the cyclophilin B group. There are many cyclophilins described, all showing sequence similarities and many characterized by their ability to bind to the immunomodulatory drug



cyclosporin A (CsA). Most cyclophilins possess enzyme activity, being peptidyl prolyl cis-trans isomerases.

The term "round spermatid protein" (RSP) is used herein to define a family of proteins sharing sequence similarity and/or exhibiting homology with the rat RSP-29 protein (described by Oneda and Djakiew, 1993, *Molecular and Cellular Endocrinology* 93, 53-61) which is a protein of rat round spermatids. RSP-29 has homologues in many different species, including higher eukaryotes (monkey to chicken), lower eukaryotes (Dictyostelium discoideum) and even prokaryotes (Rhodopseudomonas blastica and Buchnera aphidicola). The round spermatid proteins of the invention may therefore control cell differentiation in the testes as well as in the bone, and may constitute a hitherto unrecognized family of factors involved in cell differentiation (including the differentiation of bone cells such as osteoclast progenitor cells).

The term "family" is used herein to indicate a group of proteins or genes which share substantial sequence similarities, either at the level of the primary sequence of the proteins themselves, or at the level of the DNA encoding them. The sequence similarities may extend over the entire protein/gene, or may be limited to particular regions or domains. Similarities may be based on nucleotide/amino acid sequence identity as well as similarity (for example, those skilled in the art recognize certain amino acids as similar, and identify substitutions of similar amino acids as conservative changes). Some members of a protein family may be related in the sense that they share a common evolutionary ancestry, and such related proteins are herein referred to as "homologues". The members of a protein family may not necessarily share the same biochemical properties or biological functions, though their similarities are often reflected in common functional features (such as effector binding sites and substrates). The criteria by which such families are recognized are well-known in the art, and include computer analysis of large collections of sequences at the level of DNA and protein as well as biochemical techniques such as hybridization analysis.

Without being bound by theory, we propose that the bisphosphonate binding domain of the binding proteins of the invention may be characterized by the amino acid sequence motifs RGD and/or SEQ. ID. NO. 34. While other binding domains may occur, preferably, the bisphosphonate binding protein of the invention comprises the amino acid sequence RGD and/or SEQ. ID. NO. 34. Other binding proteins of the invention may have different binding domains characterized by different amino acid sequence motifs. For example, the binding protein of the invention may bind pyridoxal phosphate, O-phosphorylethanolamine, O-

phosphorylcholine, phosphatidyl ethanolamine or phospholipid bisphosphonate analogues, and/or comprise the C-terminal region of human DP1 (hDP1) and/or the amino acid sequence RGD and/or the amino acid sequence SEQ. ID. NO. 34 and/or a pyridoxal phosphate-binding domain/pseudodomain and/or a phosphatidyl ethanolamine-binding domain/pseudodomain, and/or be a Dictyostelium discoideum bisphosphonate binding protein (or homologue thereof).

Dictyostelium discoideum bisphosphonate binding proteins are particularly preferred because the present inventors have found that this organism expresses bisphosphonate binding proteins while being relatively easy to grow in large quantities. Some Dictyostelium discoideum bisphosphonate binding proteins have homologues in other organisms (for example mammalian, e.g. human cells), and such homologues are also covered by the invention.

Also contemplated by the invention are fragments of the protein of the invention which comprise the bisphosphonate binding domain, for example, comprising the amino acid sequence RGD and/or SEQ. ID. NO. 34 the C-terminal region of human DP1 (hDP1), a pyridoxal phosphate-binding domain/pseudodomain and/or a phosphatidyl ethanolamine-binding domain/pseudodomain, or any fragment substantially retains bisphosphonate binding activity.

The term "pseudodomain" is used herein to indicate that the domain is structurally and functionally related to its correlate domain, but binds a different (although usually structurally related) ligand.

Preferably, such fragments consist essentially of any bisphosphonate binding domain in the protein, preferably the bisphosphonate binding domain described above. The fragments may be fused to other peptide sequences preferably having, for example, enzyme activity or to antibodies (such as monoclonal antibodies or antibody fragments).

The bisphosphonate binding domain may be identified by any of a variety of methods known to those skilled in the art such as sequence analysis, protection analysis, affinity labeling (e.g. photoaffinity labeling) and/or the generation of a collection of fragments via e.g. protease treatment. Such techniques are known in the art

One particularly preferred bisphosphonate binding protein is DP1 (SEQ. ID NO. 3), or homologues, fragments, muteins, equivalents or derivatives (e.g. fusion derivatives or synthetic peptides) thereof which substantially retain bisphosphonate binding activity.

Another particularly preferred bisphosphonate binding protein is hDP1 (SEQ. ID NO. 9), or homologues, fragments, muteins, equivalents or derivatives

(e.g. fusion peptide derivatives or synthetic peptides) thereof which substantially retain bisphosphonate binding activity.

Yet another particularly preferred bisphosphonate binding protein is a second Dictyostelium discoideum cyclophilin herein designated DdCyP2 (SEQ. ID NO. 13), or homologues (for example, human cyclophilin B), fragments, muteins, equivalents or derivatives (e.g. fusion derivatives or synthetic peptides) thereof which substantially retain bisphosphonate binding activity.

As used herein, the term "homologue" defines proteins which are related in the evolutionary sense to the proteins of the invention, and may for example define the equivalent protein from a different organism.

The term "mutein" is used herein to define proteins that are mutant forms of the proteins of the invention, i.e. proteins in which amino acids have been added, deleted or substituted.

The term "equivalent" as used herein and applied to the proteins of the invention defines proteins which exhibit substantially the same functions as those of the proteins of the invention while differing in structure (i.e. amino-acid sequence). Such equivalents may be generated for example by identifying sequences of functional importance, selecting an amino acid sequence on that basis and then synthesizing a peptide based on the selected amino acid sequence. Such synthesis can be achieved by any of many different methods known in the art, including solid phase peptide synthesis and the assembly (and subsequent cloning) of oligonucleotides.

The term "derivative" as applied herein to the binding proteins of the invention is used to define proteins which are modified versions of the binding proteins of the invention. Such derivatives may include fusion proteins, in which the proteins of the invention have been fused to one or more different proteins or peptides (for example an antibody or a protein domain conferring a biochemical activity, such as enzymic or conjugative activity, to act as a label, or to facilitate purification).

The homologues, fragments, muteins, equivalents or derivatives of the proteins of the invention may cross-react with antibodies to the proteins of the invention, and in particular may cross-react with antibodies directed against any or all of the proteins DP1, hDP1 or DdCyP2.

In another aspect, the invention contemplates a method for producing a bisphosphonate binding protein. Preferably such a method comprises the steps of: (a) linking bisphosphonate to a chromatography column to produce an affinity column; (b) loading the binding protein (for example in the form of a cell extract)

onto the affinity column such that it becomes bound thereto; and (c) selectively eluting the binding protein from the affinity column. Methods of preparation of affinity chromatography materials are known in the art.

The step of selectively eluting the binding protein need not result in the elution of the binding protein alone in completely pure form, but merely results in a net purification (in most circumstances, a substantial net purification) of the binding protein. Selective elution as applied in the method of the invention is a process by which the binding protein is more or less specifically eluted from the column, accompanied by varying amounts of contaminating proteins and/or other (macro)molecules. Preferably, the selective elution results in the binding protein eluting as the (or a) major protein component. Preferably, the elution conditions are such that the binding protein is washed off the column substantially free of all other proteins.

The affinity chromatography material preferably comprises a bisphosphonate linked to the column. The bisphosphonate preferably comprises one or more bisphosphonate species covalently linked to the column, for example *via* a primary amine or sulphydryl group. However, any linkage may be used, so long as the linked bisphosphonate can function as a ligand in the process of affinity chromatography.

The cell extract may be a cytosolic, nuclear or membrane fraction, and can be prepared by any of a large number of techniques known to those skilled in the art. Such techniques include sonication, enzymatic lysis, centrifugation, precipitation, osmotic rupture or mechanical rupture. The cell extract is preferably a cytosolic extract, but may optionally include other cellular fractions.

Preferably, the cell from which the extract is prepared is one which is sensitive to bisphosphonate. Examples of such cells include mammalian or amoebal (e.g. *Entamoeba* spp. or *Dictyostelium* spp.) cells. Particularly preferred are cells of *Dictyostelium discoideum*. Cell lines (for example mammalian, e.g. human) expressing or overexpressing the bisphosphonate binding protein of the present invention are also useful as a source of binding protein for purification.

The binding protein may be selectively eluted from the column by loading an excess of unlinked bisphosphonate onto the column. The unlinked bisphosphonate may be different from the bisphosphonate linked to the column, and may for example be an osteoactive bisphosphonate.

The unlinked bisphosphonate for elution is preferably selected from AHBuBP, 3PHEBP, a monophosphonate analogue of 3PHEBP or PLP.

The bisphosphonate linked to the chromatography column in step (a) is preferably an osteoactive bisphosphonate. Particularly preferred are AHBuBP (Alendronate) and AHPBP (Pamidronate).

In another aspect, the invention contemplates a method for producing a bisphosphonate binding protein, comprising the steps of: (a) providing a bisphosphonate-resistant Dictyostelium mutant (e.g. a Dictyostelium discoideum mutant) bearing a mutated bisphosphonate binding protein gene; (b) cloning the wild-type gene corresponding to that mutated in step (a) to produce a cloned gene encoding a bisphosphonate-binding protein, and either: (i) expressing the cloned gene encoding a bisphosphonate binding protein to produce the bisphosphonate binding protein, or (ii) preparing a probe which is selectively hybridizable to the cloned gene and identifying a further gene on the basis of its selective hybridization with the probe and expressing said further gene.

In the case where a probe is used to identify a gene encoding a bisphosphonate binding protein, this gene may be a heterologous gene, i.e. a corresponding gene from a different biological source, a mutant thereof, etc..

The mutant for use in the above-described method is conveniently provided by the step of continuously culturing wild-type Dictyostelium discoideum Ax-2 amoebae in the presence of bisphosphonate (e.g. AHBuBP), the bisphosphonate being present at a concentration sufficient to substantially prevent growth but not sufficient to cause immediate lysis (for example, at a concentration of from 50-100  $\mu$ M for AHBuBP).

The invention as described above does not rely only on the generation of spontaneous mutants, and any of a wide variety of known mutagenesis techniques could also be used (for example those involving the treatment of the amoebae with various mutagens).

The invention also embraces bisphosphonate binding protein obtainable by the methods of the invention, as well as bisphosphonate binding proteins which have been obtained by the various methods of the invention.

In another aspect, the invention relates to isolated DNA encoding the bisphosphonate binding protein (or derivatives, fragments, and the like) thereof.

As used herein and applied to DNA, the term "isolated" indicates that the DNA is substantially isolated with respect to the complex cellular milieu in which it naturally occurs, or simply present in a different nucleic acid sequence context from that in which it occurs in nature (for example, when cloned or in the form of a restriction fragment). Thus, the DNA of the invention may be isolated in the sense

used herein, yet present in any of a wide variety of vectors and in any of a wide variety of host cells (or other milieu, such as buffers, viruses or cellular extracts).

The DNA of the present invention embraces DNA having any sequence so long as it encodes the bisphosphonate binding protein, fragment, mutein, etc. of the invention. As a result of degeneracy in the genetic code, any particular amino acid sequence may be encoded by many different DNA sequences. Determining whether a sequence encodes the above material is well within the scope of the skilled artisan as is the designing of the DNA or RNA sequence given the amino acid sequence.

In a particular embodiment, the isolated DNA of the invention has the sequence shown in Fig. 3 (SEQ. ID NO. 1), Fig. 13 (SEQ. ID NO. 9) or Fig. 21 (SEQ. ID NO. 12).

The invention also contemplates a recombinant expression vector comprising the DNA of the invention. The nature of the vector is not critical to the invention, and any vector may be used, including plasmid, virus, bacteriophage, transposon, minichromosome, liposome or mechanical carrier. As used herein, recombinant expression vector refers to a DNA construct used to express DNA which encodes a desired protein and which includes a transcriptional subunit comprising an assembly of 1) genetic elements having a regulatory role in gene expression, for example, promoters and/or enhancers, 2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and 3) appropriate transcription and translation, initiation and termination sequences. Using methodology well known in the art, recombinant expression vectors of the present invention can be constructed. Possible vectors for use in the present invention include, but are not limited to: for mammalian cells, pJT4 (discussed further below), pcDNA-1 (Invitrogen, San Diego, CA) and pSV-SPORT 1 (Gibco-BRL, Gaithersburg, MD); for insect cells, pBlueBac III or pBlueBacHis baculovirus vectors (Invitrogen, San Diego, CA); and for bacterial cells, pET-3 (Novagen, Madison, WI). The DNA sequence coding for the bisphosphonate binding protein of the invention can be present in the vector operably linked to regulatory elements.

The vector may preferably comprise an expression element or elements operably linked to the DNA of the invention to provide for expression thereof at suitable levels. Any of a wide variety of expression elements may be used, and the expression element or elements may for example be selected from promoters, enhancers, ribosome binding sites, operators and activating sequences. Such expression elements may be regulatable, for example inducible e.g., via the addition of an inducer.

As used herein, "operably linked" refers to a condition in which portions of a linear DNA sequence are capable of influencing the activity of other portions of the same linear DNA sequence. For example, DNA for a signal peptide (secretory leader) is operably linked to DNA for a polypeptide if it is expressed as a precursor which participates in the secretion of the polypeptide; a promoter is operably linked to a coding sequence if it controls the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to permit translation. Generally, operably linked means contiguous and, in the case of secretory leaders, contiguous in reading frame.

The vector of the invention can also be a viral vector, being for example based on simian virus 40, adenoviruses (e.g. human adenoviruses), retroviruses, and papillomavirus.

The vector may further comprise a positive selectable marker and/or a negative selectable marker. The use of a positive selectable marker facilitates the selection and/or identification of cells containing the vector.

Also contemplated by the invention is a host cell comprising the vector of the invention. Any suitable host cell may be used, including prokaryotic host cells (such as Escherichia coli and Bacillus subtilis), eukaryotic host cells (including mammalian cells, amoebal cells and yeast cells). Host cells may be stably transfected or transiently transfected within a recombinant expression plasmid or infected by a recombinant virus vector. Other host cells include permanent cell lines derived from insects such as Sf-9 and Sf-21, and permanent mammalian cell lines such as Chinese hamster ovary (CHO) and SV40-transformed African green monkey kidney cells (COS).

The invention also embraces a method for producing a bisphosphonate binding protein comprising the steps of: (a) culturing the host cell of the invention such that the bisphosphonate binding protein is expressed, and (b) isolating the bisphosphonate binding protein expressed in step (a).

Such recombinantly-produced bisphosphonate binding protein (recombinant bisphosphonate binding protein) can be produced relatively inexpensively in large quantities, and can be relatively easily purified, using the method of the invention.

In another aspect, the invention also contemplates a method for producing a bisphosphonate binding protein comprising the steps of: (a) probing a gene library with a nucleic acid probe which is selectively hybridizable with the DNA of the invention (for example having a sequence which is comprised in a gene corresponding to any of the sequences shown in Fig. 3 (SEQ. ID NO. 1), Fig. 13 (SEQ. ID NO. 9) or Fig. 21 (SEQ. ID NO. 12), to produce a signal which identifies a

gene that selectively hybridizes to the probe, and (b) expressing the gene identified in step (a) (for example by cloning into a host cell) to produce bisphosphonate binding protein.

As used herein, the term "selectively hybridizable" indicates that the sequence of the probe is such that binding to a unique (or small class) of target sequences can be obtained under more or less stringent hybridization conditions. This method of the invention is not dependent on any particular hybridization conditions, which can readily be determined by the skilled worker (e.g. by routine methods or on the basis of thermodynamic considerations).

The invention also embraces a bisphosphonate binding protein obtainable by the above-described methods and other methods.

The invention also relates to a nucleic acid probe which is selectively hybridizable with the DNA of the invention. Again, the term "selectively hybridizable" indicates that the sequence of the probe is such that binding to a unique (or small class) of target sequences can be obtained under more or less stringent hybridization conditions.

Preferably, the nucleic acid probe is selectively hybridizable with (for example having a sequence which is comprised in) a gene corresponding to any of the sequences shown in Fig. 3 (SEQ. ID NO. 1), Fig. 13 (SEQ. ID NO. 9) or Fig. 21 (SEQ. ID NO. 12).

The protein of the invention may be labeled, for example with a fluorescent label, an antibody, a radioisotope or an enzyme. Such labeled proteins may be particularly suitable for use in the test kits of the invention.

The invention also covers various uses of the bisphosphonate binding proteins of the invention. For example, the invention contemplates a method for screening for therapeutically active bisphosphonates, new and potentially active bisphosphonates and their analogues, as well as entirely new classes, not broadly defined as bisphosphonates, comprising the steps of: (a) contacting a bisphosphonate with the bisphosphonate binding protein of the invention, and (b) determining whether binding between the bisphosphonate and binding protein occurs, wherein binding is indicative of a therapeutically active bisphosphonate.

The method described above is useful for screening large numbers of bisphosphonates and bisphosphonate like compounds for therapeutic activity. It may also be used to classify or differentiate potential osteoactive or anti-arthritis bisphosphonates according to their mode of action (e.g., by their cellular targets). In fact, it may be used in ranking efficacy of the compounds to be screened. Preferably, the method is employed in high throughput screening. Compounds



identified by the method of the invention can be further modified or used directly as therapeutic compounds to activate or inhibit the natural functions of the binding protein encoded by the isolated DNA molecule, for example in the treatment of osteoporosis.

The invention also provides a method for evaluating the therapeutic activity of a bisphosphonate comprising the steps of: (a) contacting the bisphosphonate with a bisphosphonate binding protein of the invention, and (b) measuring the binding affinity of the bisphosphonate binding protein for the bisphosphonate.

For example, the method described above is useful for ranking the therapeutic activity of potential bisphosphonate drugs, bisphosphonate analogs, and "bisphosphonate mimetics" (acting in a similar manner to a bisphosphonate) drugs.

In another aspect, the invention provides a method for synthesizing a therapeutically active bisphosphonate comprising the steps of: (a) generating a three-dimensional model of the bisphosphonate binding site of the bisphosphonate binding protein of the invention, and (b) modeling the therapeutically active bisphosphonate with reference to the three-dimensional model generated in step (a).

Many different techniques exist for generating a three-dimensional model of the binding protein (or fragments/derivatives thereof) for use in the above-described methods, and all are suitable for use in the method of the invention. Conveniently, the three-dimensional model is generated by computer analysis of the amino-acid sequence of all or a portion of the bisphosphonate binding protein (for example the bisphosphonate binding domain). Alternatively, the three-dimensional model could be generated by X ray crystallography or by NMR of the binding protein (or fragments/derivatives thereof). These techniques could also be applied to the bisphosphonate binding protein-bisphosphonate complex, the results of which could also be used as the basis for the rational design of therapeutic agents.

The invention also embraces a therapeutically active bisphosphonate which has been screened, evaluated or synthesized by the methods described above.

Also contemplated by the present invention are test kits comprising the bisphosphonate binding protein of the invention. Such kits are useful, for example, in the screening and evaluating methods of the invention. The bisphosphonate binding proteins comprised in the kits of the invention are preferably bound to a solid support and may conveniently include a labeled (e.g. radioactively-labeled, fluorescently labeled, enzymatically labeled) bisphosphonate (for example for use in competitive binding assays or in displacement assays).

Also contemplated by the invention are antibodies which bind to the binding protein of the invention. Such antibodies can be prepared by employing standard

techniques well known to those skilled in the art, using any of the bisphosphonate binding proteins of the invention as antigens for antibody production. These antibodies can be employed in assays, diagnostic applications, therapeutic applications, and the like. Preferably, and particularly for therapeutic applications, the antibodies are monoclonal antibodies.

The antibodies of the invention may advantageously bind specifically to the bisphosphonate binding proteins of the invention. Antibodies specific for the bisphosphonate binding site may act as bisphosphonate mimetics. Specific binding may be exploited in imaging techniques, for example to assess the extent to which bisphosphonate targets are available for bisphosphonate action, or to determine the degree of occupancy of bisphosphonate targets in patients undergoing bisphosphonate therapy. They may also be used to identify and isolate new bisphosphonate binding proteins.

The invention also contemplates antibody derivatives, including antibody fragments (e.g. Fab fragments), chimaeric antibodies (including humanized antibodies) and antibody derivatives (such as fusion derivatives comprising an antibody-derived variable region and a non-immunoglobulin peptide having for example enzyme or conjugative activity).

The invention also contemplates mimetics (for example, the antibodies of the invention described above) or antagonists of the bisphosphonate binding protein of the invention.

The bisphosphonate binding proteins of the invention find utility in a wide range of therapeutic applications, and in another aspect the invention contemplates the bisphosphonate binding protein, antibody thereto, mimetic or antagonist thereof for use in therapy.

The bisphosphonate binding proteins, antibodies thereto, mimetics or antagonists thereof can be administered in a clinical setting by intraperitoneal, intramuscular, intravenous, or subcutaneous injection, implant or transdermal modes of administration, and the like.

Such administration can be expected to provide therapeutic alteration of the activity of bone active agents, and in preferred embodiments, the therapy involves: (i) the regulation of bone metabolism, for example in the treatment of Paget's disease, hypercalcaemia (both tumour-induced and non-tumour induced), bone metastases and osteoporosis, or (ii) the regulation of sperm maturation, for example for contraception or fertility treatment, or (iii) the regulation of bone metabolism via interaction with cyclosporin A, and related compounds that bind to immunophilin proteins.

Knowledge of the hDP1 gene sequence disclosed herein and its expression may be useful for understanding the spermatogenic process. In testis, germ cells are surrounded by Sertoli cells and both cell types interact physically and chemically via a unique and impressive array of structural features. Since hDP1 appears to be one of the factors from round spermatids that modulate the Sertoli cell function, it provides a basis for the development of novel contraceptives, or agents that affect reproduction in other ways, such as infertility, by affecting sperm maturation.

The recognition of the role of cyclophilins in bisphosphonate binding and as targets for osteoactive bisphosphonate provides a new therapeutic use for cyclosporins (e.g. cyclosporin A), and related compounds as modulators of bone metabolism, in the treatment of bone metabolism disorders and as adjuncts to bisphosphonate treatments.

#### EXAMPLES

The invention is described in further detail by way of specific non-limiting examples. These examples are purely exemplary and are not intended to be limiting in any way. The skilled artisan will appreciate that these examples may be varied and used as guidance, in light of the art, to make and use the invention as claimed. The examples may refer to drawings and descriptions offered previously.

##### Example 1: Isolation of DPL, a bisphosphonate-binding discoidin from Dictyostelium Materials

The di- or tri-sodium salts of AHPrBP (Pamidronate) and AHBuBP (Alendronate) were from GENTILI S.P.A., Pisa, Italy; 3PHEBP (Risedronate) is obtained from Procter and Gamble Pharmaceuticals, Cincinnati, OH, USA, <sup>32</sup>P-dCTP is purchased from Amersham. Unless stated otherwise, all other chemicals were from Sigma Chemical Co., Poole, UK.

##### Construction of an immobilized bisphosphonate affinity column

Two bisphosphonate-affinity columns were prepared. One prepared using AHBuBP and the other using AHPrBP, were made using the same column matrix and the same coupling procedure. AHBuBP and AHPrBP each have a primary amino group available for the coupling to the matrix. Spectra/Gel MAS Beads (Spectrum, Los Angeles, CA, USA) were used as the matrix to which an aldehyde group is attached through a five atom hydrophilic spacer arm. The bisphosphonates were coupled to the aldehyde via their amino groups in the presence of sodium cyanoborohydride (NaCNBH<sub>4</sub>) as described in detail below.

A 1 M stock solution of sodium cyanoborohydride (NaCNBH<sub>4</sub>) is prepared and left at room temperature for 1 to 3 hours until the bubbling subsided before use. The gel is washed on a Buchner funnel with three volumes of coupling buffer (0.1 M

phosphate buffer, pH 7.8). AHBuBP or AHPPrBP is dissolved to a high concentration (400 mM for AHBuBP and 150 mM for AHPPrBP) in the coupling buffer and added to the same volume of the gel. Sodium cyanoborohydride from the stock solution is added to a final concentration of 0.1 M and the suspension is agitated at room temperature overnight. The coupled gel is agitated with a small amount of sodium borohydride ( $\text{NaBH}_4$ ) for 2 hours at room temperature in order to reduce any unreacted aldehyde groups. The gel is then washed with 10 volumes of 0.5 M NaCl, followed by column equilibration buffer (30 mM MOPS, 0.1 mM EDTA and 0.5 mM DTT, pH 7.4). 15 ml of the gel were packed into a glass column and stored in the equilibration buffer containing 0.02% (w/v) sodium azide at 4°C.

After each use, the column is regenerated by washing with a minimum of 10 volumes of 0.5 M NaCl, 0.1 Tris-HCl, pH 8.5, followed by a minimum of 10 volumes of 0.5 M NaCl, 0.1 M Na acetate, pH 4.5. The column is cleared of any irreversibly bound protein by washing with 5 M urea. The column is stored at 4°C in the equilibration buffer containing 0.02% (w/v) sodium azide.

#### Growth and harvest of Dictyostelium discoideum cells

Dictyostelium discoideum amoebae strain Ax-2, were grown as shaken cultures in HL5-glucose medium at 22°C (Watts and Ashworth, *Biochem. J.* 119: 171-174, 1970). Cells from two 700 ml cultures were harvested at a density of  $5\text{--}10 \times 10^6$  cells/ml and disrupted by sonication in 20 ml 30 mM MOPS, 0.1 mM EDTA, 0.5 mM DTT, pH 7.4 at 4°C. Cell debris is removed by ultracentrifugation at 184,000g at 4°C for 2 hours.

#### Purification of bisphosphonate-binding protein DPl

The supernatant, containing cytosolic proteins, is then loaded onto a pre-equilibrated affinity column at a rate of 6-12 ml/hour. The column is washed first with equilibration buffer at a rate of 20 ml/hour until protein, monitored by absorbance at 280 nm, no longer appeared in the eluate. The ionic strength of the eluant is then increased by including 0.1 M KCl in the equilibration buffer in order to wash away further non-specifically bound proteins. Finally, specific elution of proteins bound to immobilized bisphosphonate is achieved by eluting with buffer containing 5 mM bisphosphonate. In order to examine whether other compounds were also able to elute proteins which were still bound to the column following a 0.1M KCl wash, 5 mM inorganic phosphate, 5 mM pyrophosphate, 5 mM PLP or 300 mM D-galactose were added to the elution buffer. All procedures were carried out at 4°C.

Fractions eluted from the columns were lyophilised, resuspended in 2-3 ml distilled water, and dialysed against three changes of 50 mM ammonium

bicarbonate. The dialysed samples were further concentrated approximately 50-fold by ultrafiltration in Microsep microconcentrators (Filtron, Northborough MA, USA) with a 10kDa cut-off filter. The concentrated samples were examined by SDS-PAGE analysis on 12% polyacrylamide gels with standard molecular weight markers (200kD, 97kD, 68kD, 43kD, 28kD and 18kD) (Gibco, Paisley, Scotland) and stained with Coomassie brilliant blue R-250 (Laemmli, *Nature* 227, 680-685, 1970).

A typical elution profile, following loading of cell-free extract onto the immobilized bisphosphonate affinity columns is shown in Fig. 1. A cell extract prepared by sonication of *Dictyostelium* cells in equilibration buffer at 4°C is loaded onto an AHBuBP-affinity column at a rate of 10 ml/hour. Fractions (each represented by a chromatography monitor peak) of non-specifically-bound proteins were eluted with equilibration buffer (peak 1), equilibration buffer after overnight incubation (peak 2) and equilibration buffer containing 0.1 M KCl (peak 3), respectively (as described above). DP1 is eluted when 5 mM AHBuBP dissolved in equilibration buffer is used directly (peak 4), after overnight incubation (peak 5) and after further incubation for a few hours (peak 6).

Each peak from the column is lyophilised, desalted by dialysis and concentrated in a Microsep microconcentrator. An aliquot from each fraction is electrophoresed on a 12% polyacrylamide gel in denaturing conditions (Laemmli, *Nature* 227, 680-685, 1970) and the gel is stained with Coomassie Blue. Elution with equilibration buffer, equilibration buffer after overnight incubation and with buffer containing 0.1 M KCl eluted proteins which had no affinity for the immobilized ligand (Fig. 2, Lanes 2, 3 and 4 respectively). Subsequent washing with 5 mM bisphosphonate (either AHBuBP or AHPPrBP) resulted in the elution of DP1 alone, of about 28kDa (Fig. 2, lane 5). Lane 1 contains molecular weight standards.

DP1 binds to both the AHBuBP and the AHPPrBP affinity columns very strongly, since repeated overnight incubation of the columns with 5 mM bisphosphonate followed by elution the following day continued to elute DP1 over a period of 10 days.

DP1 is the only protein eluted from the affinity-column by bisphosphonates since it is the only protein that could be detected in a 3,000-fold concentrate of the bisphosphonate eluate by SDS-PAGE analysis and staining with Coomassie blue. The high purity is later confirmed by matching all of the fragments obtained by peptide sequencing of protein in the bisphosphonate eluate with the predicted amino acid sequence of the DP1 cDNA.

Not all bisphosphonates elute DP1 from the columns. Whereas AHBuBP, AHPuBP and 3PHEBP at a concentration of 5 mM appear to be equally effective, 5 mM HEBP, a bisphosphonate which is 12-fold less potent at inhibiting *Dictyostelium* growth than AHBuBP, is not able to elute DP1 at all. This demonstrates that DP1 appears to bind specifically to second and third generation, potent anti-resorptive bisphosphonates.

Neither pyrophosphate, inorganic phosphate nor galactose is effective at eluting DP1. However, pyridoxal 5'-phosphate (PLP), at a concentration of 5 mM, eluted DP1 very effectively, although some DP1 is still retained on the column which could be eluted with 5 mM AHBuBP. Additionally, a number of other proteins were also eluted by PLP.

#### Enzymatic cleavage and peptide sequencing

The purified DP1 protein is digested with endoprotease Lys-C or endoprotease Asp-N in the presence of urea as denaturing agent to produce a collection of peptides for sequencing. These peptides were separated by HPLC and sequenced by Edman degradation (Edman, *Acta Chem. Scand.* 4: 283-293, 1950).

Nine peptide sequences were determined by Edman degradation (Table I). A search in data banks for similar sequences showed that peptides 2, 3, 4, 5, 6, 8 and 9 had considerable identity with peptides from discoidin I, a lectin isolated from *Dictyostelium discoideum*. By contrast, peptides 1, and 7 had little homology with any peptides in discoidin I. This implies that DP1 is related to discoidin I but is not discoidin I itself.

**Table 1. Peptide Sequences from DP1**

Peptide No.	Produced by endo-protease	Peptide Sequence	% identity with sequences also found in discoidin I
1 (SEQ. ID NO. 14)	Lys-C	NSILNFSNSK	
2 (SEQ. ID NO. 15)	Lys-C	HFV(AorN)ISTQGRGDHDQXVTXY	55%
3 (SEQ. ID NO. 16)	Lys-C	GTGsRTIV	50%
4 (SEQ. ID NO. 17)	Lys-C	DASRFDGSWSSXVLDK	
5 (SEQ. ID NO. 18)	Lys-C	LRYTLDNVNWVEYNNGEINANK	
6 (SEQ. ID NO. 19)	Lys-C	XRSIAIHPOTYNNHIsIr	79%
7 (SEQ. ID NO. 20)	Asp-N	dNGQMRWEGKSENI	
8 (SEQ. ID NO. 22)	Asp-N	DLTFITWGNNAVY	54%
9 (SEQ. ID NO. 23)*	Lys-C and Asp-N	DSVKHFVAISTQGRGDHDQWVTSY KLRYTLDNVNWVEYNNGEIINANK	52%

\*FIG. 3 (peptide 9 is a deduced sequence since peptides isolated after Asp-N digestion of DP1 showed that peptides 2 and 5 are contiguous).

Endoproteinase Lys-C cleaves at the carboxyl side of a lysine peptide bond and Asp-N cleaves at the amino side of an Asparagine peptide bond. Lower case letters indicate a tentative sequence assignment. "X" indicates that no definite assignment could be made at that position. Bold-typed regions are those used for generating oligonucleotides XJ-1 (peptide 5) and XJ-2 (peptide 6) in the chart.

Example 2: Cloning and sequencing of DP1Isolation of poly(A<sup>+</sup>) mRNA and synthesis of first strand cDNA

Messenger RNA is isolated from approximately 10<sup>7</sup> *Dictyostelium* amoebae using a Quick-Micro mRNA purification kit (Pharmacia, Piscataway NJ, USA). For the synthesis of first strand cDNA, 2.5 ug mRNA were precipitated with potassium acetate and ethanol, collected by centrifugation at 14,000 g for 5 minutes at 4°C, and resuspended in 50 ul DEPC-treated water. 12.5 ng oligo(dT)<sub>17</sub>-adapter primer, GACTCGAGTCGACATCGA(dT)<sub>17</sub> (SEQ. ID NO. 23), were added and the solution heated to 70°C for 10 minutes followed by quick chilling on ice for 3 minutes. One-fifth volume of 5 x reaction buffer (250 mM Tris-HCl, pH8.3, 0.375 mM KCl, 15 mM MgCl<sub>2</sub>), 80 units RNasin (Promega, Madison WI, USA), 1.25 mM dNTP and 0.2 mM DTT were added and the mixture is prewarmed at 37°C for 2 minutes, followed by addition of 600 units Superscript RNase H<sup>-</sup> reverse transcriptase (Gibco, Paisley, UK) and incubation at 37°C for 1 hour. The cDNA is diluted and used for PCR.

Amplification of cDNA by 3'-RACE

Sequences from peptides 5 and 6 described above were used to design the degenerate oligonucleotide primers XJ-1 and XJ-2. Alignment of the peptides from DP1 with the sequence of discoidin I suggested that peptide 5 would be closer than peptide 6 to the N-terminal end of DP1. Hence XJ-1 is expected to be 5' of XJ-2 in the mRNA sequence.

The cDNA coding for DP1 is amplified by 3'-RACE (Frohman et al. *Proc. Natl. Acad. Sci. USA* 85, 8998-9002, 1988) using degenerate oligonucleotide primers:

(a) XJ-1 (SEQ. ID NO. 5):

[TA(T/C)ACI(T/C)TIGA(T/C)AA(T/C)GTIAA(T/C)TGGGT]

(b) XJ-2 (SEQ. ID NO. 6):

[(C/A)G(T/A/G)(T/A)(C/G)TAT(T/C/A)GCIATICA(T/C)CC]

in sequential PCRs. In the first round the reaction included 50 mM KCl, 10 mM Tris-Cl, pH9.0, 0.1% Triton X-100, 1.5 mM MgCl<sub>2</sub>, 0.2 mM dNTP, diluted cDNA, 0.2 M XJ-1 and XJ-ADA [GACTCGAGTCGACATCGA] (SEQ. ID NO. 24) primers, and 2.5 units Taq DNA polymerase (Promega, Madison, WI, USA). Each cycle is initiated by denaturation at 94°C for 1.5 minutes, and continued with annealing at 46°C for 1.5 minutes and extension at 72°C for 2 minutes. After 40 cycles, samples were incubated for a further 15 minutes at 72°C and then stored at 4°C. The resulting PCR products were diluted 1:1,000 with distilled water and used as the template for the second-round PCR reaction with the XJ-2 and XJ-ADA

primers. Reaction conditions were identical to those used in the first round. PCR products were size fractionated on a 1% agarose gel, visualized and isolated (Sambrook J., Fritsch E. F., and Maniatis T. (eds): *Molecular Cloning: A laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989). The PCR product is cloned into the pCRII vector (Invitrogen, San Diego CA, USA) according to the manufacturer's instructions. Those recombinant plasmids that contained inserts were sequenced (Sanger et al., *Proc. Natl. Acad. Sci. USA* 74: 5463-5467, 1977).

The primer that is used for reverse transcription of mRNA consisted of an oligo(dT)<sub>17</sub> region to anneal to the Poly(A) tail of the mRNA and an adapter region of 18 nucleotides (Frohman et al., *Proc. Natl. Acad. Sci. USA* 85, 8998-9002, 1988). Incorporation of the adapter sequence into the first-strand cDNA allowed use of the primer XJ-ADA for the amplification of the 3' region of the DP1 cDNA using PCR.

Initially, PCR is performed using XJ-1 and XJ-ADA as 5' and 3' primers respectively to amplify the first-strand cDNA. Two DNA fragments, with approximate sizes of 600bp and 500bp were detected on subsequent agarose gel electrophoresis. Sequencing of these two fragments showed that neither of them encoded any of the peptide sequences obtained from DP1. When primers XJ-2 and XJ-ADA were used to amplify the first-strand cDNA, one band of about 400bp is observed on an agarose gel. Sequencing of this band indicated that this cDNA is a fragment from the 3' end of discoidin IA cDNA. This is not surprising since primer XJ-2 is derived from a peptide sequence for which there is a homologous sequence in discoidin I. However, when, in a second round of the PCR, the products of the reaction using primers XJ-1 and XJ-ADA were further amplified using primers XJ-2 and XJ-ADA, a DNA fragment of 450bp is obtained. Cloning and sequencing of this fragment showed that its corresponding amino acid sequence contained the sequences of peptides 3, 6, 7 and 8 of DP1 (Fig. 3 (SEQ. ID NO. 2) and Fig. 4).

This DNA fragment, designated XJ-450, is used as a probe for the subsequent cDNA library screening and northern and Southern blot analyses.

#### Preparation and screening of cDNA libraries

Poly(A<sup>+</sup>) RNA isolated from the wild-type Ax-2 strain of *Dictyostelium* is used as the template for cDNA synthesis for construction of libraries in lambda ZAP bacteriophage (Stratagene, San Diego, CA, USA). To isolate full-length clones, the libraries were screened with the PCR amplified cDNA fragment (XJ-450) of DP1 isolated from the pCRII vector. Hybridisations were performed at 42°C for 24 to 48 h in 5 x SSPE - 1 x Denhardt's solution - 100 ug salmon testis DNA/ml - 50% formamide. Membranes were washed in 0.25 x SSPE, 0.5% SDS at 55°C.



Rescreening of positive plaques yielded full-length cDNAs for DP1 from the wild-type Ax-2 library.

When  $1.2 \times 10^6$  plaques from the cDNA library of the wild-type Ax2 strain of *D. discoideum* were screened with XJ-450, nine positive clones were obtained. Analysis of the sequence of the cDNAs (Fig. 3) (SEQ. ID NO. 1) revealed that the inserts comprise 968 bp containing an open reading frame of 771 bp which codes for 257 amino acids. The untranslated regions on each side of the coding region are rich in A+T and the 3' untranslated region also contains multiple consensus polyadenylation sequences (AATAAA) (Fig. 3) (SEQ. ID NO. 1).

#### Northern and Southern analysis of DP1

For northern analysis, 0.5 ug of poly(A<sup>+</sup>) RNA isolated from *Dictyostelium* amoebae is size fractionated on a denaturing formaldehyde agarose gel, transferred onto Hybond-N membrane (Amersham, Bucks, UK), hybridized to the appropriate cDNA fragment and washed as described above. To control for variations in loading and transfer of RNA, the northern blots were probed with a fragment of Dd-tcp1, a *Dictyostelium* homologue of the tcp1 family of molecular chaperones.

Genomic DNA is isolated from the wild-type Ax-2 strain by using a modification of the methods described by Nellen et al. (Molecular Biology in *Dictyostelium*: Tools and Applications. In: Spudich JA ed: *Methods in Cell Biology* Vol. 28: 67-100. Academic Press, 1987) and Arnau (*Laboratory Products Technology* / September 1993:20.).  $2 \times 10^9$  *Dictyostelium* amoebae were harvested, washed with distilled water at 4°C and lysed in ice-cold HMN (0.01 M Mg acetate, 0.01 M NaCl, 0.03 M HEPES and 10% sucrose, pH 7.5) containing 0.5% NP-40. Nuclei were separated from the cell lysate by centrifugation at 10,000g for 10 min. at 4°C, resuspended in 1 ml extraction buffer (0.4 M KCl, 0.05 M EDTA, 1% Triton-X100) and treated with 3 µl RNase I (7.5 u/µl) (Promega, Madison, WI, USA) at 70°C for 15 min. Insoluble material is removed by centrifugation at 10000g for 2 min. at room temperature and the DNA in the supernatant is purified (Magic Miniprep DNA Purification System, Promega, Madison, WI, USA). Purified DNA is quantified by absorbance at OD<sub>260</sub> and 10 µg samples were digested overnight with the required restriction enzymes. The DNA is then size fractionated on a 1% agarose gel, transferred to Hybond-N membrane and hybridized and washed as described above. A Southern blot of genomic DNA prepared by Clontech (Palo Alto, CA, USA) is hybridized with the XJ-450 DP1 DNA probe using low stringency for hybridization (37°C) and washing (3 x SSPE, 0.5% SDS).

Northern analysis of Poly(A<sup>+</sup>) RNA isolated from amoebae that had been grown either axenically to a density of about  $5 \times 10^6$  cells/ml or to a lower density of  $1 \times 10^5$  cells/ml, or with a bacterial substrate is shown in Fig. 5a. The probe hybridized to a mRNA transcript of about 1.2 kb from axenic Ax-2 amoebae harvested at a density of about  $5 \times 10^6$  cells/ml. The expression of DP1 mRNA in wild-type cells grown with a bacterial food source appeared to be very low, although the presence of the mRNA is confirmed by repeating the northern blot with a larger amount of poly(A<sup>+</sup>) RNA. DP1 mRNA is also expressed in axenic wild-type amoebae harvested at a lower density of  $1 \times 10^5$  cells/ml (Fig. 5b). For comparison, a PCR fragment of discoidin IA is labeled with <sup>32</sup>P-dCTP and hybridized to the same northern blots. Like DP1, discoidin I is also expressed poorly in bacterially-grown wild-type Ax-2 amoebae. (Fig. 5a).

Southern analysis of genomic DNA revealed that only a single band from each restriction enzyme digest hybridized to the XJ-450 probe for DP1 (Fig. 6b). On the same blot, a discoidin IA probe hybridized to at least three bands in each DNA digest (Fig. 6c).

This is consistent with the claim that discoidin I is encoded by a family of genes (Rowekamp et al., *Cell* 20: 495-505, 1980; Poole et al., *J. Mol. Biol.* 153:273-289, 1981) but suggests that there is only a single gene for DP1 (discoidin II).

#### Structural features of DP1

DNA and protein sequences were compared with updated releases from GenBank, Swissprot and Entrez, and analyzed by using the computer software MacVector (International Biotechnologies, Inc.), GCG (The Wisconsin Genetic Computer Group (Devereux et al., *Nucleic Acids Res.* 12: 387-395, 1984) CLUSTALV (Higgins et al., *Comp. Appl. Biosci.* 8: 189-191, 1992), and MacProt (Markiewicz, *BioTechniques* 10: 756-757+760+762-763, 1991) 1991; Luttke, *Comp. Method. Prog. Biomed.* 31:105-112, 1990 ).

The predicted amino acid sequence suggests that DP1 is a neutral protein with an estimated pI of 6.77. Its calculated molecular mass is 28,573Da which is slightly larger than that of discoidin IA (28,258Da) or discoidin IC (28,391Da) (Poole et al., *J. Mol. Biol.* 153: 273-289, 1981). The amino acid sequence predicted from the full-length cDNA includes all nine peptide sequences obtained by enzymatic digestion of DP1. The amino acid and cDNA sequences of DP1 were later found to be identical to those of discoidin II. The homology between discoidin II and I is significant, with 49% identity in their predicted amino acid sequences (Fig. 4) and 61% identity in the nucleotide sequences of their open reading frames.

In addition to regions of identity with discoidin I, DP1 (discoidin II) shares partial homologies with several other proteins, including human coagulation factors V (Jenny et al., *Proc. Natl. Acad. Sci. USA* 84:4846-4850, 1987) and VIII (Wood et al., *Nature* 312:330-337, 1984.), milk fat globule protein (Larocca et al., *Cancer Res.* 51: 4994-4998, 1991) and open reading frame 7 linked to the Rhodopseudomonas blastica *atp* operon (Tybulewicz et al., *J. Mol. Biol.* 179: 185-214, 1984) (Fig. 7). Hydropathicity analysis by the Kyte-Doolittle (*J. Mol. Biol.* 157, 105-132, 1982) or the Hopp-Woods (*Proc. Natl. Acad. Sci. USA* 78: 3824-3828, 1981) algorithms showed that the first 20 amino acids at the amino terminus of DP1 are hydrophobic (Fig. 8) and a signal peptide cleavage site, predicted by using the method of von Heijne (*Nucleic Acids Res.* 14: 4683-4690, 1986), appears to be present between two serines at residues 20 and 21 (Fig. 9). Secondary structure predictions by using the methods of Chou-Fasman (*Adv. Enzymol. Relat. Areas Mol. Biol.* 47, 45-148, 1978) and Garnier-Robson (*J. Mol. Biol.* 120, 97-120, 1978) suggest that DP1 is mainly composed of  $\beta$ -pleated sheets (Fig. 8). An Arg-Gly-Asp (RGD) motif occurs at residues 81-83 and is also present in discoidin I. The amino acids flanking the RGD motif are also highly conserved between discoidin I. and DP1. Owing to its hydrophilicity, the RGD motif probably lies on the protein surface (Fig. 10). The motif may have the same role in DP1 and in discoidin I.

#### Example 3: Recombinant Expression of DP1

In order to conveniently produce large quantities of DP1, the DNA encoding it (see Fig. 3 (SEQ. ID NO. 1)) is transferred into an appropriate expression vector and introduced into mammalian cells by standard genetic engineering techniques well-known to those skilled in the art.

Suitable vectors include pCD (Okayama et al., *Mol. Cell. Biol.* 2: 161-170 (1982) and pJL3/4 (Gough et al. *EMBO J.*, 4: 645-653 (1985).

The resultant expression vectors are then transformed into appropriate host cells, which thereby express the DP1 protein.

Those skilled in the art could also manipulate the sequence shown in Fig. 3 (SEQ. ID NO. 2) to introduce mammalian regulatory sequences and/or expression signals. Alternatively, it could be manipulated to introduce bacterial expression signals, and the resultant recombinant DNA could then be introduced into bacterial host cells (for example Escherichia coli) for expression therein.

Similar manipulations could be performed for the construction of insect vectors and yeast vectors.

#### Expression in mammalian cells

The cDNA inserts of the pBluescript SK(-) subclones described in Example 2 (above) are subcloned into the mammalian cell expression vector pMT2 using appropriate restriction enzymes followed by ligation. The structure of the resultant expression vector is confirmed by restriction mapping on agarose gels.

Plasmid DNA from the pMT2 subclone is then transfected into monkey COS-1 cells by the DEAE-dextran procedure (Sompayrac and Danna (1981) *Proc. Natl. Acad. Sci. U.S.A.*, 78: 7575-7578; Luthman and Magnusson (1983), *Nucl. Acids Res.* 11: 1295-1308). Serum-free 24 hour conditioned medium is then collected from the cells starting 40 to 70 hours post-transfection.

#### Example 4: Isolation of hDP1, a human bisphosphonate binding protein

The tri-sodium salts of AHBuBP (Alendronate) is from GENTILI Materials S.P.A., Pisa, Italy. Human multiple tissue blots and a zoo-blot were purchased from Clontech. Human poly(A<sup>+</sup>) RNA from testis, HeLa cells and HL60 cells is purchased from Clontech. A cDNA library of human testis is custom-made in lambda ZAP bacteriophage by Stratagene (La Jolla, CA, USA) using poly(A<sup>+</sup>) RNA purchased from Clontech (Palo Alto, CA, USA). The human leukocyte concentrate is obtained from Sheffield Blood Transfusion Services (Longley Lane, Sheffield, UK). All other chemicals were purchased from Sigma Chemical Co., Poole, UK.

#### Construction of a bisphosphonate affinity column

A bisphosphonate-affinity column is constructed as described previously in Example 1.

#### Preparation of human leukocytes

$3-4 \times 10^9$  human white blood cells were washed twice with phosphate buffered saline (PBS; pH 7.4) and disrupted by sonication in 40 ml of column equilibration buffer. Insoluble material is removed by ultracentrifugation at 184,000 x g at 4°C for 2.0 hours to yield a supernatant.

#### Purification of hDP1

The supernatant containing the cytosolic proteins is applied to the pre-equilibrated AHBuBP-affinity column at a rate of 6-12 ml/hour. The column is first eluted with equilibration buffer at a rate of 20 ml/hour until protein no longer appeared in the eluate. The ionic strength of the eluant is then increased by including 0.1 M KCl to elute further non-specifically bound proteins until protein no longer appeared in the eluate. Finally, buffer containing 5 mM AHBuBP specifically eluted proteins bound to the immobilized AHBuBP on the column. Fractions eluted from the column were lyophilised and resuspended in 2-3 ml distilled water, followed by extensive dialysis against 50 mM ammonium bicarbonate. The dialysed samples were further concentrated to 50-100 µl by

ultrafiltration using Microsep microconcentrators (Filtron, Northborough MA, USA) having a 10kDa cut-off filter. Proteins in each concentrate were separated by SDS-PAGE (Laemmli, *Nature* 227, 680-685, 1970) and stained with Coomassie brilliant blue R-250. The 28kDa protein isolated from the AHBuBP-affinity chromatography is hDP1 (SEQ. ID NO. 13).

#### Protein sequencing

N-terminal amino acid sequencing of hDP1 (SEQ. ID NO. 13) is performed on a concentrated solution of the AHBuBP eluate and also on the band of hDP1 (SEQ. ID NO. 13) transferred onto PVDF membrane by protein blotting following SDS-PAGE. Sequence is determined by Edman degradation (Edman, *Acta Chem. Scand.* 4: 283-293, 1950) both directly from sample solution and from the (PVDF) membrane containing blotted protein. Internal peptides were generated by endoprotease Lys-C digestion, purified by HPLC, and sequenced.

Both approaches yielded the same N-terminal sequence:

MKVEVLPALTDNYMYLVIDDETKEAAIVDPVQ (SEQ. ID NO. 25) [peptide 10]

Cleavage of hDP1 by endoprotease Lys-C produced an internal peptide having the sequence:

YXIGEPTVPsTLAEeFtYNpF (SEQ. ID NO. 26) [peptide 11]

where X indicates that no definite assignment is made for that residue, and a lower case letter indicates a tentative assignment for that residue.

The N-terminal amino acid sequence (peptide 10) (SEQ. ID NO. 25) is highly homologous to the N-terminal sequence of a rat spermatid (SEQ. ID NO. 8) 29,000 MW protein (Figure 11). No protein appeared to contain sequences having significant homology with peptide 11 (SEQ. ID NO. 26).

#### Example 5: Cloning and sequencing of hDP1

##### Synthesis of degenerate oligonucleotides and RT-PCR

Since the N-terminal sequence of hDP1 showed homology to a rat spermatid protein, poly(A<sup>+</sup>) RNA of human testis is used to prepare total cDNA that contained the hDP1 cDNA.

Degenerate oligonucleotides containing 17-22 nucleotides were synthesized according to the peptide sequences of hDP1 for the subsequent PCR to amplify the corresponding hDP1 cDNA (Table 2). Synthesis of the first strand cDNA from human testis poly(A<sup>+</sup>) mRNA is carried out using a 3' RACE kit (Gibco, Paisley, UK) following the manufacturer's instructions. AP-oligo(dT) is used as the primer since this introduced the sequence of the universal amplification primer (UAP) into

the cDNA sequence to facilitate the subsequent 3' RACE procedure (Frohman et al., *Proc. Natl. Acad. Sci. USA* 85, 8998-9002, 1988).

Table 2. Nucleotide sequences of primers used for the PCR reaction

Oligonucleotides	Directions	Nucleotide sequences (5'-3')
DJ1 (SEQ. ID NO. 27)	sense	GCIYTIACNGAYAAAYTAYATGTA
DJ5 (SEQ. ID NO. 28)	sense	GAYGAYGARACNAARGARGC
DJ9 (SEQ. ID NO. 29)	sense	ATHGGIGARCCACIGTIGG
DJ10 (SEQ. ID NO. 30)	antisense	TADCCICTYGGITGICAIGG
AP (SEQ. ID NO. 31)	antisense	GGCCACGCGTCGACTAGTAC(T) <sub>17</sub>
UAP (SEQ. ID NO. 32)	antisense	CUACUACUACUAGGCCACGCGTCGACTAGTAC

"I" indicates inosine

"N" indicates either C, A, G or T

"Y" indicates either C or T

"R" indicates either G or A

The strategy for the amplification of hDP1 (SEQ. ID NO. 15) cDNA is shown in Fig. 12. Initially, the PCR is performed in a volume of 50  $\mu$ l in the presence of AmpliTaq reaction buffer, 1.5 mM MgCl<sub>2</sub>, 0.2 mM dNTP, diluted cDNA template, 2.5 units of AmpliTaq DNA polymerase (Perkin Elmer) and 0.2  $\mu$ M of each primer DJ-1 (SEQ. ID NO. 27) and UAP (SEQ. ID NO. 32) (Table 2). The mixture is subjected to 40 cycles of amplification in a Perkin Elmer 9600 PCR machine. Each cycle involved four successive steps including denaturation at 94°C for 15 sec, annealing at 48°C and 60°C for 15 sec, and extension at 72°C for 15 sec. At the end of cycling the sample is further incubated for 10 min. at 72°C and then kept at 4°C. Second-round PCR is carried out by using the diluted reactions of the first-round as template and different combinations of internal primers, i.e. DJ-1 (SEQ. ID NO. 27) plus DJ-10 (SEQ. ID NO. 30), DJ-5 (SEQ. ID NO. 28) plus DJ-10, DJ-9 (SEQ. ID NO. 29) plus UAP (SEQ. ID NO. 32) (Table 2). 10  $\mu$ l of each reaction is run into a 1% low melting point agarose gel (Gibco, Paisley, UK). The desired fragment is purified using the Wizard PCR DNA purification system (Promega, Madison, WI, USA) and the yield is quantified.

Use of primers DJ-1, which encoded part of peptide 1, and UAP (Table 2) in the polymerase chain reaction gave DNA products that formed a faint smear on an agarose gel. This DNA is reused as the template for further amplifications with internal primers. In the second round of the PCR, DNA fragments were amplified by two combinations of primers: (1) hA : a fragment of 600bp from DJ-1 and DJ-10; (2) hB : a fragment of 400bp from DJ-9 and UAP. Fragments hA and hB, which were expected to be contiguous with each other (Fig. 12), added up to a size of 1,000bp. It is found that when hA is used as the template for PCR amplification with primers DJ-5 and DJ-10, a fragment (hC) of 570 bp is obtained. All these sequences were found within the full length sequence (SEQ. ID NO. 9). All these

three fragments were from specific amplifications since leaving out either primer of a pair, or the template, eliminated the corresponding product. hC is used as the probe for screening the cDNA library and for the subsequent northern and Southern blotting.

#### cDNA library screening and sequence of the full-length of hDP1 cDNA

25 ng of a hDP1 PCR fragment is labeled with  $^{32}\text{P}$ -dCTP using random primers (Feinberg and Vogelstein, *Anal. Biochem.* 132: 6-13, 1983) (Prime-It II kit, Stratagene) and is used as a probe to screen a human testis cDNA library in lambda ZAP bacteriophage (Short et al., *Nucleic Acid Res.* 16: 7853-7860, 1988). After three rounds of screening, bacteriophage plaques specifically hybridized to the hDP1 probe were selected and the inserts were subcloned into pBluescript SK(-) plasmids by *in vivo* excision (Short et al., *Nucleic Acid Res.* 16: 7853-7860, 1988). The recombinant plasmids were subjected to restriction digestion and amplification using T3 and T7 primers which anneal to sequences on each side of the inserts. Recombinant plasmids, which contained a 1.2 kb hDP1 cDNA insert, were identified by Southern blotting and sequenced using a Taq DyeDeoxy Terminator Cycle Sequencing kit (ABI) by the method of Sanger *et al.* (*Proc. Natl. Acad. Sci. USA* 74: 5463-5467, 1977).

When hC is used to screen a testis cDNA library constructed in lambda ZAP bacteriophage, three clones were isolated. All appeared to have the same nucleotide sequence. The largest open reading frame consisted of 936 bp and contained six ATG codons. The N-terminal sequence of hDP1 corresponded with the deduced amino acid sequence from the cDNA only if the second ATG codon of the cDNA encoded the N-terminal methionine of hDP1. The cDNA nucleotide sequence then indicated that hDP1 consisted of 260 amino acids and had a calculated molecular mass of 28,872 and an estimated isoelectric point of 7.0 (Fig. 13). The predicted molecular weight is in good agreement with that of the protein observed on a polyacrylamide gel.

If translation started from the first ATG codon, an additional polypeptide of 48 amino acids would be produced at the N-terminus of hDP1 (Fig. 13) (SEQ. ID NO. 9) to give a protein of 34 kDa. This putative peptide is designated hDP1-rP. The full-length cDNA contains a 5' untranslated region of 157 nucleotides and a 3'-untranslated region of 265 nucleotides. However, neither a poly(A) tail nor a classical AATAAA polyadenylation signal is apparent in the cDNA towards the 3' end, probably owing to restriction digestion beyond the polyadenylation site in construction of the cDNA library.

#### Human tissue-type distribution and inter-species conservation of the hDP1 gene

A  $^{32}\text{P}$ -dCTP labeled hDP1 cDNA fragment is hybridized to human multiple tissue northern blots (Clontech) and to a northern blot containing mRNA isolated from cultured human cells THP1 (a monocytic cell line), MG63 (an osteoblast-like cell line) and primary bone cells. In addition, the hDP1 probe is also hybridized to a northern blot which contained mRNA from Dictyostelium discoideum and to a zoolot which contained genomic DNA from different species (Clontech, Palo Alto, CA, USA).

Hybridization is performed at 42°C in the presence of 50% formamide (Sambrook *et al.*, *Molecular Cloning: A laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989). Washing at high stringency is carried out at 42°C using 3 x 100 ml 0.25 x SSPE plus 0.5 % SDS for 3 x 15 min. A 2kb DNA fragment of human b-actin (Clontech, Palo Alto, CA, USA) is also labeled with  $^{32}\text{P}$ -dCTP and is used as a positive control for hybridization. The hybridization signal is quantified with a BioRad Phosphorimager.

DNA is also prepared from a lambda gt11 cDNA library of normal human osteoclast-like multinucleated cells using the Lambda DNA Purification Kit (Stratagene, Cambridge, UK). After being digested by restriction enzymes, the bacteriophage DNA is separated on a 1% agarose gel, transferred to a Hybond-N membrane and hybridized to the radiolabeled hDP1 probe.

Expression of hDP1 in several human tissues is studied by northern analysis with the PCR fragment hC of hDP1 cDNA as a probe. The hybridized bands, corresponding to 1.2 kb mRNA transcripts, suggest that the cDNA clone obtained for hDP1 is almost full-length, even though it did not contain the region having the polyadenylation signal. Moreover, the hDP1 cDNA fragment hybridized to mRNA from all the human tissues and cell lines studied (Fig. 14). These comprised heart, brain, placenta, liver, lung, skeletal muscle, kidney, pancreas, spleen, thymus, prostate, testis, ovary, small intestine, colon, peripheral blood leukocytes and human cell lines HeLa, HL60, THP-1, MG63 and primary bone cells. Among all these tissues and cells, testis, skeletal muscle and heart had the highest levels of hDP1 mRNA expression. On the other hand, placenta, pancreas, spleen and peripheral blood leukocytes appear to be the tissues that contain least hDP1 mRNA. In addition, a hybridized band is also observed on the Southern blot of DNA prepared from a cDNA library of human osteoclast-like multinucleated cells, suggesting the presence of the hDP1 mRNA transcripts in osteoclasts.

Southern blots of *Eco*RI digests of genomic DNA from various organisms showed that hybridization at high stringency with the hC fragment produced one or more hybridized bands in all species studied except yeast (Fig. 15). In addition, a



hybridized band corresponding to a mRNA transcript with a size of 1.8 kb is also detected in a northern blot containing mRNA isolated from Dictyostelium amoebae growing under various conditions (Fig. 16).

Computer assisted analyses on the DNA and protein sequences of hDP1

DNA and protein sequences were compared with the updated releases from GenBank, Swissprot and Entrez, and analyzed by using the computer software MacVector (International Biotechnologies, Inc.), GCG (The Wisconsin Genetic Computer Group) (Devereux *et al.*, *Nucleic Acids Res.* 12: 387-395, 1984), CLUSTALV (Higgins *et al.*, *Comp. Appl. Biosci.* 8: 189-191, 1992) and FRODO (Jones, *Method Enzymol.* 115: 157-171, 1985).

The N-terminal amino acid sequence of hDP1 shared striking homology with a rat 29 kDa round spermatid protein in which the N-terminal 32 amino acids had been determined by peptide sequencing (Fig. 11) (SEQ. ID NO. 8) (Onoda and Djakiew, *Mol. Cell. Endocrinol.* 93: 53-61, 1993). A search through data banks with the complete deduced amino acid sequence of hDP1 showed that there is no exact match to any known sequence. However, there is significant homology with the amino acid sequences predicted from the DNA sequences of two genes from prokaryotic organisms: open reading frame 3 linked to the Rhodopseudomonas blastica *atp* operon (ATPase-ORF3) (33% identity over the entire sequence) (Tybulewicz *et al.*, *J. Mol. Biol.* 179, 185-214, 1984) (Fig. 17(a)) and open reading frame 1 linked to the genes of arginyl tRNA synthetase and ribonuclease H of Buchnera aphidicola (ATS-ORF1) (30% identity over the entire sequence) (Munson *et al.*, *Gene* 137: 171-178, 1993) (Fig. 17(b)).

Among the proteins with known functions, the aspartate transaminase (AAT) of chicken mitochondria (14% identity and 44% similarity) showed some homology to hDP1 (Jaussi *et al.*, *J. Biol. Chem.* 260:16060-16063, 1985) (Fig. 18). However, AAT is larger than hDP1 and the alignment of hDP1 with the three-dimensional structure of AAT suggests that most of the conserved amino acid residues are probably in the internal region of the folded protein. It is interesting that residues around the PLP binding site are fairly well conserved although lysine does not exist in the same position in hDP1 as in AAT.

A hydropathicity plot using the Kyte-Doolittle (Kyte and Doolittle, *J. Mol. Biol.* 157, 105-132) or the Hopp-Woods (Hopp and Woods, *Proc. Natl. Acad. Sci. USA* 78: 3824-3828, 1981) algorithms shows that the C-terminal part of the hDP1 protein is quite hydrophilic with a high surface probability (Fig. 19). Use of the methods of Chou-Fasman (*Adv. Enzymol. Relat. Areas Mol. Biol.* 47, 45-148, 1978) and Garnier-Robson (Garnier *et al.*, 1978, *J. Mol. Biol.* 120, 97-120) to predict



secondary structure suggests that this region mainly comprises  $\alpha$ -helix. No obvious signal peptide cleavage site is predicted by the method of von Heijne (*Nucleic Acids Res.* 14: 4683-4690, 1986).

A search in the data bank with the sequence of the putative N-terminal peptide, hDP1-rP, showed no significant homology to any existing sequences. Prediction by the method of von Heijne (*Nucleic Acids Res.* 14: 4683-4690, 1986) suggested that there is a possible signal peptide cleavage site between alanine and arginine residues at positions -21 and -22 (Fig. 13). The scoring of this prediction, however, is not sufficiently high to allow a firm conclusion.

Several potential protein motifs have been suggested for hDP1 by searching the database of protein motifs. These include a site of  $\text{Ca}^{2+}$ -calmodulin dependent kinase II at residues 228-232 (REKTV) near the C-terminus and a part of a zinc-finger at residues 53-61 (LTTHH).

#### Sequence comparison of hDP1 and DP1

The protein sequence of hDP1 is also compared with that of DP1. No striking homology is observed between the two proteins (with 13% identity and 38% similarity). One region of identity is seen between residues 44-47 (His-Gly-Val-Lys) (SEQ. ID NO. 10) of hDP1 and residues 28-31 of DP1 (SEQ ID NO. 3). This sequence is hydrophilic and is located at the surface of hDP1 protein (Fig. 20).

#### Example 6: Recombinant Expression of hDP1

Large quantities of hDP1 can be obtained using procedures similar to those described above in Example 3.

#### Expression in mammalian cells

The cDNA inserts of the pBluescript SK(-) subclones described in Example 5 (above) were subcloned into the mammalian cell expression vector pMT2 using appropriate restriction enzymes followed by ligation. The structure of the resultant expression vector is confirmed by restriction mapping on agarose gels.

Plasmid DNA from the pMT2 subclone is then transfected into monkey COS-1 cells by the DEAE-dextran procedure (Sompayrac and Danna (1981) *PNAS*, 78: 7575-7578; Luthman and Magnusson (1983), *Nucl. Acids Res.* 11: 1295-1308). Serum-free 24 hr conditioned medium is then collected from the cells starting 40 to 70 hours post-transfection.

#### Example 7: Isolation of DdCyP2, a bisphosphonate binding cyclophilin from Dictyostelium

AHPrBP, AHBuBP and Cl<sub>2</sub>MBP were from GENTILI S.P.A., Pisa, Italy. All other bisphosphonates and the monophosphonate (NE10788) were from Procter

and Gamble Pharmaceuticals, Cincinnati, OHIO, USA. PLP is from Sigma (Poole, Dorset, UK).

#### Construction of a bisphosphonate affinity column

Two bisphosphonate-affinity columns were prepared as described previously in example 1.

#### Growth and harvest of Dictyostelium discoideum cells

Dictyostelium discoideum strain Ax-2 is grown in HL-5 glucose medium (2 x 700 ml) on a shaker at 22°C to a density of  $2-8 \times 10^6$  cells/ml. Cells were harvested and washed once with distilled water and once with equilibration buffer at 4°C. The cell pellet is resuspended in 3-4 volumes of equilibration buffer and disrupted by sonication for 4 x 20 seconds in a MSE Soniprep with an interval of 1 minute on ice between each sonication. The cell lysate is centrifuged at 45,000 rpm for 2.0 hours at 4°C in a Beckman 50.2 Ti rotor.

#### Isolation of the bisphosphonate-binding proteins from Dictyostelium

The supernatant, containing the cytosolic proteins is loaded onto the AHBuBP-affinity column, which had been pre-equilibrated with the equilibration buffer, at about 6-12 ml/hour. After loading, the column is washed with the equilibration buffer at a rate of 20 ml/hour to wash off unbound proteins and then washed with 0.1 M KCl in the equilibration buffer to wash off further non-specifically bound proteins. After such washing, bisphosphonates or PLP at a concentration of 5 mM in equilibration buffer were used to elute any bisphosphonate-binding proteins. The eluates (about 25 ml) containing the proteins were lyophilised overnight and redissolved in 2-3 ml of 10 mM ammonium formate. They were dialysed against four changes of 2500 ml ammonium formate over 48 hours. The dialysed proteins were further concentrated by centrifugation in Microsep microconcentrators with a molecular weight cut-off of 10 kDa (Filtron, Northborough, MA, USA) at 5,000 g for 1-3 hours. All purification and concentration processes were carried out at 4°C.

Proteins in each concentrated fraction were separated by electrophoresis in denaturing conditions on a 12% polyacrylamide gel. Molecular weight markers (Gibco, Paisley, Scotland) were also run on the gel. The gels were stained either with silver or Coomassie blue.

PLP eluted two major proteins with molecular weights of 18 kDa and 28 kDa respectively. The 18 kDa protein (which is present in approximately the same amount as the 28 kDa protein) is later named DdCyP2 (SEQ. ID NO. 13). The 28 kDa protein is probably DP1.

#### Peptide sequencing and peptide sequence analysis

Following electrophoresis, the gel is equilibrated for 5 minutes in the transfer buffer [10 mM CAPS, 10% (v/v) methanol, pH 11.0] prior to blotting to remove electrophoresis buffer salts and detergents. The membrane is cut to the dimensions of the gel, wetted with 100% methanol, and then equilibrated with the transfer buffer for 15-30 minutes. Two filter papers (Whatman 3 MM) and fibre pads were completely saturated in the transfer buffer. The gel, membrane and pads were assembled into the holder cassette. The transfer is performed at constant voltage (50 V) for 1 hour at room temperature with the Bio-Ice cooling unit in the buffer chamber.

Proteins blotted onto ProBlott membrane were detected by Coomassie brilliant blue staining following the instructions supplied with the membrane. The blotted membrane is rinsed with deionised water and saturated with 100% methanol. The membrane is then stained with 0.1% (w/v) Coomassie blue R-250 in 40% (v/v) methanol, 1% (v/v) acetic acid. Protein bands appeared within one minute. The membrane is destained with 50% (v/v) methanol, and rinsed extensively with deionised water and air dried. The bands of interest were excised for protein sequencing.

The N-terminal sequence of DdCyP2 is determined three times directly after blotting onto ProBlott membrane using repeated cycles of Edman degradation. The consensus sequence obtained is: NH<sub>2</sub> - GKDPKITNKV FFDE (SEQ ID NO. 33).

#### AHPrBP-affinity column

An AHPrBP-affinity column is also used for isolation of bisphosphonate-binding proteins by the same purification procedure as that used with the AHBuBP-affinity column. AHBuBP and PLP both eluted only the 28 kDa protein off this column. No DdCyP2 could be detected in either the PLP fraction or the AHBuBP fraction on a SDS-PAGE gel. Experiments confirmed that DdCyP2 is never eluted from the AHPrBP column, even with 5 M urea. It is probable that DdCyP2 does not bind to the AHPrBP-affinity column.

#### Homology analysis

DNA and protein sequences were extracted from the updated releases from the GenBank or Swissprot. The sequences were analyzed using the program GCG (The Wisconsin Genetic Computer Group) (Devereux et al., *Nucleic Acids Res.* 12: 387-395, 1984).

Homology analyses of the amino acid sequence with the protein database suggested DdCyP2 had strong homology to cyclophilins, especially cyclophilin Bs. The molecular weight of cyclophilins is usually 18-22 kDa, which is very similar to that of DdCyP2.

#### Example 8: Cloning and sequencing of DdCyP2

DdCyP2 (SEQ. ID NO. 12) is cloned and sequenced from a Dictyostelium discoideum strain Ax-2 cDNA library, using similar methods to those described in Example 2 (above). The nucleotide and predicted amino acid sequence are shown in Fig. 21 (SEQ. ID NO.12 and SEQ. ID NO.13).

The amino acid sequence of DdCyP2 is compared with that of the other bisphosphonate-binding protein DP1 of Dictyostelium. There is little similarity in their amino acid sequences (17% amino acid identity with four gaps), except that they both contain the RGD motif.

Amino acids 85-87 in DdCyP2 form an Arg-Gly-Asp (RGD) tripeptide motif. This RGD motif is also present in the sequences of all known animal cyclophilin Bs (CyPBs), including human, mouse, rat and chick but not in other members of the cyclophilin family (e.g. CyPA). CyPBs are a type of CyP occurring in the endoplasmic reticulum (ER). The CyPB family has a N-terminal signal sequence, which directs it to the ER, and a conserved C-terminal undecapeptide extension which is not found in other CyPs. A comparison of the amino acid sequences of DdCyP2 and the CyPBs from various species showed about 60% identity with one gap introduced into the CyPBs. On the basis of the "RGD" motif, DdCyP2 is like a CyPB, but it does not have the extra N- and C-terminal sequences typical of CyPBs. DdCyP2 is therefore like a hybrid between CyPA and CyPB; no similar CyPs have been reported in other cells.

The amino acid sequence of DdCyP2 is superimposed on that of hCypA by using the program FRODO (Jones, *Method Enzymol.* 115: 157-171, 1985). DdCyP2 is probably very similar to human cyclophilin A (hCypA) in its three-dimensional structure. The RGD tripeptide and the seven amino acid insertion in DdCyP2 both appear on the surface of the structure.

#### Example 9: Recombinant expression of DdCyP2

Large quantities of DdCyP2 can be obtained using procedures similar to those described above in Example 3.

##### Expression in mammalian cells

The cDNA for DdCyP2 described in Example 8 is subcloned into the mammalian cell expression vector pMT2 using appropriate restriction enzymes followed by ligation. The structure of the resultant expression vector is confirmed by restriction mapping on agarose gels.

Plasmid DNA from the pMT2 subclone is then transfected into monkey COS-1 cells by the DEAE-dextran procedure (Sompayrac and Danna (1981) *PNAS.* 78: pages 7575-7578; Luthman and Magnusson (1983), *Nucl. Acids Res.* 11: pages

1295-1308). Serum-free 24 hr conditioned medium is then collected from the cells starting 40 to 70 hours post-transfection.

Example 10: Production of antibodies to bisphosphonate binding proteins

Following the procedures set out in the "Guide to Protein Purification", Deutscher (Ed), Methods in Enzymology, Academic Press, pages 663-679 and "Antibodies, Volume I, a practical approach", Catty (Ed), IRL Press (the contents of which are hereby incorporated herein by reference), antibodies are produced as follows:.

A. Polyclonal antibodies

20-200 ug of a highly pure preparation of hDP1 is emulsified with Freund's complete adjuvant (FCA) using two 3 ml Luer-Lock syringes with 18-gauge needles. The emulsified immunogen mixture is then injected intradermally into a shaved rabbit using a 22-gauge needle. A total volume of 0.5-1.0 ml is injected into 10-12 sites along the upper sides of the rabbit. Boosters are given after 21 days and 40 days.

After 40 days, the rabbit is bled and the blood allowed to clot for 2-4 hours at room temperature. The serum is then decanted and centrifuged at 1000g for 10 min. to remove blood cells.

Immunoglobulin is then partially purified from the serum by DEAE-Sephacel ion-exchange chromatography. The serum is first dialyzed against phosphate buffered saline, pH 7.4, and then applied to the column. Anti-hDP1 IgG antibody is then collected as it flows off the column.

B. Monoclonal antibodies

The procedure followed is based on that described by Kohler and Milstein (1975) Nature 256: 495.

50-100 ug purified hDP1 in 0.2 ml complete Freund's adjuvant is injected into the hind foot pad of a mouse. After 10-12 days, the swollen popliteal lymph node in the hind leg is dissected from the surrounding fat pad.

The lymph node is placed into Dulbecco's minimal essential media which has been supplemented with 2 mM glutamine, 100 IU/ml penicillin and 100 ug/ml streptomycin. The tissue is minced and the cells dispersed. The cell suspension is allowed to settle for 10 min. on ice and the supernatant then centrifuged (1600 rpm for 6-7 min.). The supernatant is discarded. An equal volume of medium is then added and the cells washed twice. Viable cells are counted using dye exclusion.

The cells are then fused with the p3U1 myeloma cell line (Yelton et al. (1978) Curr. Top. Microbiol. Immunol. 81. 1), using standard techniques.

Cells (2ml/well) are then pipetted into cell culture trays and placed in an incubator. Multiple subclonings are then carried out and the hybridomas screened by ELISA.

Example 11: Bisphosphonate binding immunoassay

The assay is based on a modification of the Antigen Capture Assay (originally developed by Berson and colleagues - see Berson, S.A., Yalow, R.S., Bauman, A., Rothschild, M.A. and Newerly, K. (1956) Journal of Clinical Investigation 35, pages 170 to 190, and Yalow, R.S. and Berson, S.A. (1959) Nature 184, pages 1648-1649).

Anti-bisphosphonate binding protein monoclonal antibody is bound to the bottom of each well of a microtitre plate using standard procedures (approximately 2 ug/ml of antibody in PBS). Any remaining binding sites in the wells are then blocked with 3% BSA in PBS and a partially pure preparation of bisphosphonate binding protein is added to each well to specifically bind the bisphosphonate binding protein to the antibodies in the wells. Each well is then washed to remove excess unbound bisphosphonate binding protein and any other free contaminating proteins.

Radiolabeled (eg.  $^{14}\text{C}$  or  $^3\text{H}$ ) bisphosphonate is then introduced into the wells. The labeled bisphosphonate binds to the bisphosphonate binding protein which is attached to the antibody-coated wells. The plates are incubated for a time sufficient to allow full binding and the wells are then washed to remove unbound bisphosphonate.

The amount of radiolabel in each well is then determined by washing each well with trichloroacetic acid followed by liquid scintillation counting of the resulting fluid. The binding affinity of a range of different bisphosphonates can therefore be determined by comparing the binding characteristics of radiolabeled derivatives in the assay.

Example 12: Bisphosphonate binding competitive immunoassay

Anti-bisphosphonate binding protein monoclonal antibody is bound to the bottom of each well of a microtitre plate using standard procedures (approximately 2 ug/ml of antibody in PBS), and the microtitre plate treated as described in Example 11.

The ability of an unlabeled test bisphosphonate to compete with a labeled reference bisphosphonate for binding to the immobilized bisphosphonate binding protein is then determined by assaying a series of test mixtures having predetermined concentrations of labeled reference and unlabeled test bisphosphonates using assay procedures similar to those described in Example 11.

Example 13: Unlabeled bisphosphonate binding assay

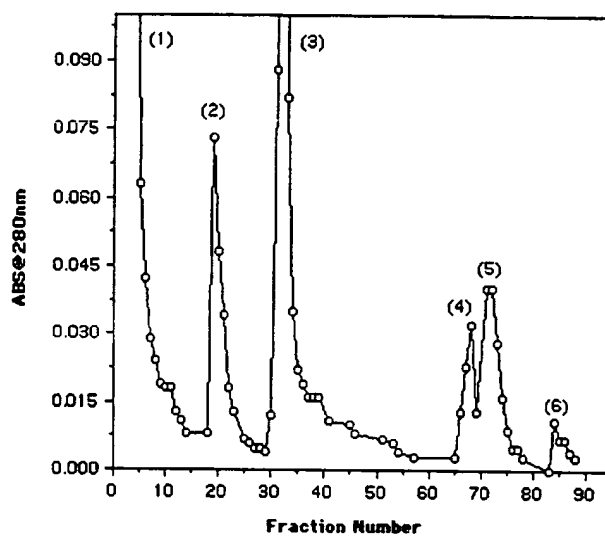


The bisphosphonate binding protein is bound to a Pharmacia BIAcore 2000 chip using standard procedures. Small volumes of bisphosphonate solutions at various concentrations are then allowed to flow over the chip, and changes in the optical properties of the chip attendant on bisphosphonate binding are detected by the Pharmacia BIAcore apparatus. The results permit calculation of the binding constant for the binding of the bisphosphonate to the bisphosphonate binding protein.

The content of all references cited herein are hereby incorporated by reference. Materials described as starting materials are made by known methods, or are known in the art. They are thus practicable to the skilled artisan without further guidance.

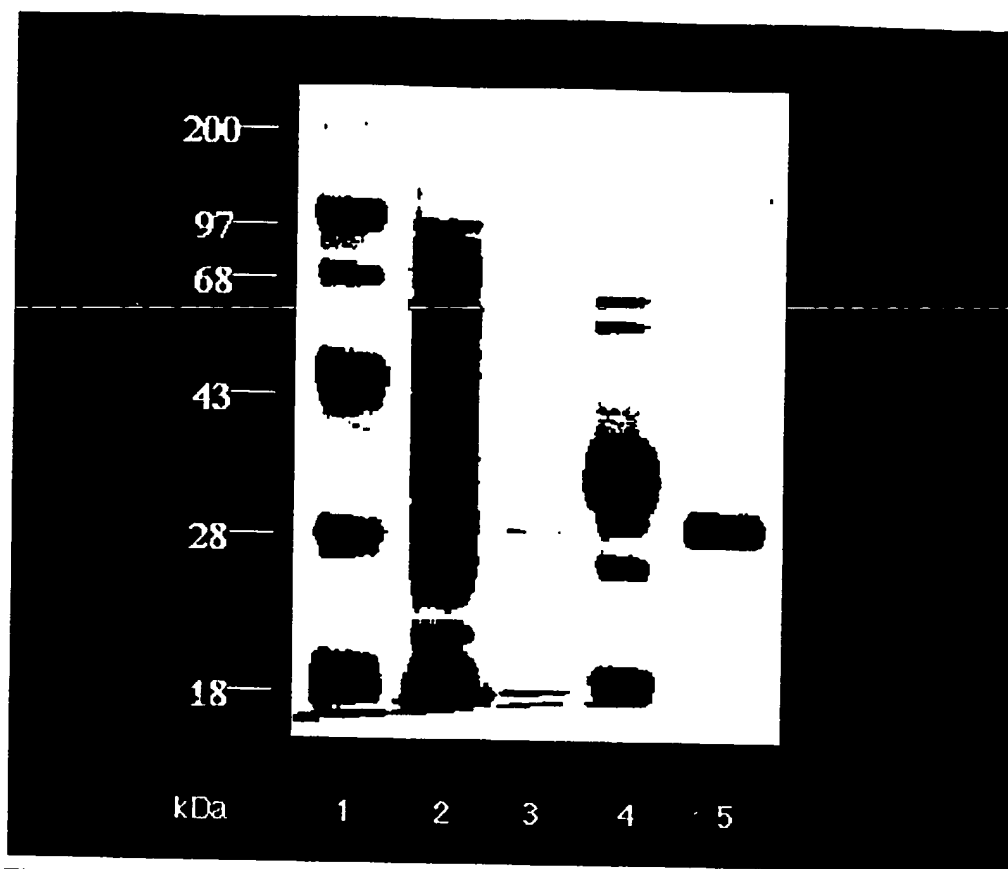
Though not required for enablement, the Dictyostelium Ax-2, referred to herein, is deposited with the ATCC, (American Type Culture Collection, Rockville, Maryland, USA) prior to submission of this application and bears deposit number 24397, see the ATCC/NIH Repository Catalogue.

Patent Figure 1



**Figure 1. Affinity chromatography of a cell extract of *Dictyostelium discoideum* on an AHBuBP-affinity column.**

The cell extract prepared by sonication of *Dictyostelium* cells in equilibration buffer at 4°C is loaded onto an AHBuBP-affinity column at a rate of 10 ml/hour. Peaks of non-specifically-bound proteins were eluted with equilibration buffer (1), equilibration buffer after overnight incubation (2) and equilibration buffer containing 0.1 M KCl (3). DP1 is eluted when 5mM AHBuBP dissolved in equilibration buffer is used directly (4), after overnight incubation (5) and after a further incubation for a few hours (6)



**Figure 2. Profile of eluted proteins from the AHBuBP-affinity column.**

Each peak from the column is lyophilised, desalted by dialysis and concentrated in a microsep microconcentrator. An aliquot of each fraction is electrophoresed on a 12% polyacrylamide gel in denaturing conditions (Laemmli, 1970) and the gel is stained with Commassie Blue. Lane 1, molecular weight standards; Lane 2, elution with equilibration buffer; Lane 3, elution with equilibration buffer after overnight incubation; Lane 4, elution with equilibration buffer containing 0.1 M KCl; lane 5, elution with 5 mM AHBuBP in equilibration buffer.

3 TTTTATTAGTATTTAATTTATTTATATAATTCTTTTAAAAAAAACAAAACAAAACAAATG  
1 M  
51 TCC GTT CCA GCT GGT TCT GTT TCA TGT CTT GCT AAT GCA TTA TTA AAT  
2 S V P A G S V S C L A N A L L N  
99 TTA AGA TCA TCA ACT GAT TAT AAT GCT GAT CAT GGT GTA AAG AAT TCT  
18 L R S S T D Y N A D H G V K N S  
147 ATT TTA AAT TTT TCA AAT TCA AAG GAT GCT AGT AGA TTC GAC GGT AGT  
34 I L N F S N S K D A S R F D G S  
195 GAA TCA TGG TCA TCA TCA GTT TTG GAT AAG AAT CAA TTC ATT GTT GCC  
50 E S W S S S V L D K N Q F I V A  
243 GGT AGT GAT TCT GTT AAA CAT TTC GTT GCA ATC TCA ACT CAA GGT CGT  
66 G S D S V K H F V A I S T Q G R  
291 GGT GAT CAT GAT CAA TGG GTA ACT TCA TAC AAA TTA AGA TAC ACA CTT  
82 G D H D Q W V T S Y K L R Y T L  
339 GAT AAT GTA AAC TGG GTT GAA TAT AAC AAT GGT GAA ATA ATC AAT GCC  
98 D N V N W V E Y N N G E I I N A  
387 AAT AAA GAT AGA AAT TCA ATT GTT ACA ATC AAC TTT AAT CCA CCA ATT  
114 N K D R N S I V T I N F N P P I  
435 AAA GCT AGA TCT ATT GCC ATT CAT CCT CAA ACC TAT AAT AAT CAT ATT  
130 K A R S I A I H P Q T Y N N H I  
483 TCA CTT CGT TGG GAA TTA TAT GCA TTA CCA GTT AAA AGT TAT TCA AAT  
146 S L R W E L Y A L P V K S Y S N  
531 CCA TCA GTC CAA GTT GGT GAA GTT TCA ATT GGT GAT AGA TCT CTT AAC  
162 P S V Q V G E V S I G D R S L N  
579 AGT GGT ACT GGT TCA CGT ACG ATT GTT CGT CAC GTT AAA TTC CCA GTG  
178 S G T G S R T I V R H V K F P V  
627 GAA TTC CTT TCT GTT CCA ATC GTA TCA ATT GGT TGT AAA AAA GTT GAT  
194 E F L S V P I V S I G C K K V D  
675 GCA CAT ACT GAT AAT GGT CAA ATG AGA TGG GAA GGT AAA TCT GAA AAT  
210 A H T D N G Q M R W E G K S E N  
723 ATT ACT ACA AAA GGT TTT GAT TTA ACT TTT ATT ACA TGG GGT AAT AAT  
226 I T T K G F D L T F I T W G N N  
771 GCA GTT TAT GAT TTA ACT TTT GAT TAT GTT GCT GTT GAA TTT AAT AAT  
242 A V Y D L T F D Y V A V E F N N  
833 TAA ATAATTAAATAATAAAATAAAATAAAATTTATTTGTTTTATTTATATTTTAAAT

\*

```

896      AATTAAATAATTAAATAATTAAAAAAAAAAAAAAAAAAAAAAAAATTTTAAATTTTCCAGAAAAA
          AAAAAAAAAAAAA
908

```

**Figure 3. Nucleotide and deduced amino acid sequence of the cDNA for wild-type DP1.**

The Nucleotide sequence shown is of the 986-bp cDNA of DP1 (Discoidin II) from the wild-type *Dictyostelium discoidium* Ax-2. The adenine base of the ATG initiation sequence is assigned as 1 in the numbering. Nucleotides are numbered in the *right margin* and amino acids on the *left*. An open reading frame of 771-bp encoding 257 amino acids is shown with a *single letter code* for the translated amino acids. A termination codon (TAA) at the end of the coding sequence is marked with an *asterisk*. A putative signal peptides at the beginning of the amino acid sequence (1-20) is indicated in *italics*. The nucleotides sequence of the PCR fragment XJ-450 used to screen the cDNA library is *underlined*. Multiple polyadenylation signal sequences (AATAAA) are shown in *italics*.

DISC IA QYIVAGCEVPRTFMCVALQGRGDADQWVTSYKIRYSLDNVSWFEYRNGAAVTGVTDRNTV  
 DP1 QFIVAGSDSVKHFVAISTOGRGDHDQWVTSYKLRYTLDNVVWVEYNNGEIINANKDRNSI  
 \*.\*\*\*\*\* . . \* . . \*\*\*\*\* \*\*\*\*\* . . . . . \*\*\*\*\* \* \* \* \* \* . . . . .  
 (Peptide 2) (Peptide 5)

DISC IA MSTQ-GLVQLLANAQCHLRTSTNYNGVHTQFNSALNYKN-NGTNTIDGSEAWCSSIVDTN  
 DP1 MSVPAGSVSCLANALLNLRSSSTDYNADHGVKNSILNFSNSKDASRFDGSESWSSSVLDKN  
 \*\* \* \* . . . . . \*\*\*\*\* \* \* . . . . . \* \* \* \* \* . . . . . \*\*\*\*\* \* \* \* \* \*  
 (Peptide 1) (Peptide 4)

DISC IA QYIVAGCEVPRTFMCVALQGRGDADQWVTSYKIRYSLDNVSWFEYRNGAAVTGVTDRNTV  
 DP1 QFIVAGSDSVKHFVAISTOGRGDHDQWVTSYKLRYTLDNVVWVEYNNGEIINANKDRNSI  
 \*.\*\*\*\*\* . . \* . . \*\*\*\*\* \*\*\*\*\* . . . . . \*\*\*\*\* \* \* \* \* \* . . . . .  
 (Peptide 2) (Peptide 5)

DISC IA VNHFFDTPIRARSAIHPLTWNGHISLRCEFYTPQVQS---SVTQVGADIYTGDNCA<sup>1</sup>LN<sup>2</sup>T  
 DP1 VTINFNPPIKARSIATHPOTYNNHISLRWELYPVKSYSNPSVQVG-EVSI<sup>3</sup>GDR-SLNS<sup>4</sup>  
 \* . . \*\*\*\*\* \* \* \*\*\*\*\* \* . . \* \* \* . . \* \* \* . . \* \* \* . .  
 (Peptide 6)

DISC IA GSGKREVVVPVKFQFEFATLPKVALNFDQIDCTDATNQTRIGVQPRNITTKGFDCVFYTW  
 DP1 GTGSRTIVRHVKFPVEFLSVPIVSIGCKKVD<sup>5</sup>AHTDNGOMRWEGKSENITTKGF<sup>6</sup>DLTE<sup>7</sup>ITW<sup>8</sup>  
 \* . \*  
 (Peptide 3) (Peptide 7) (Peptide 8)

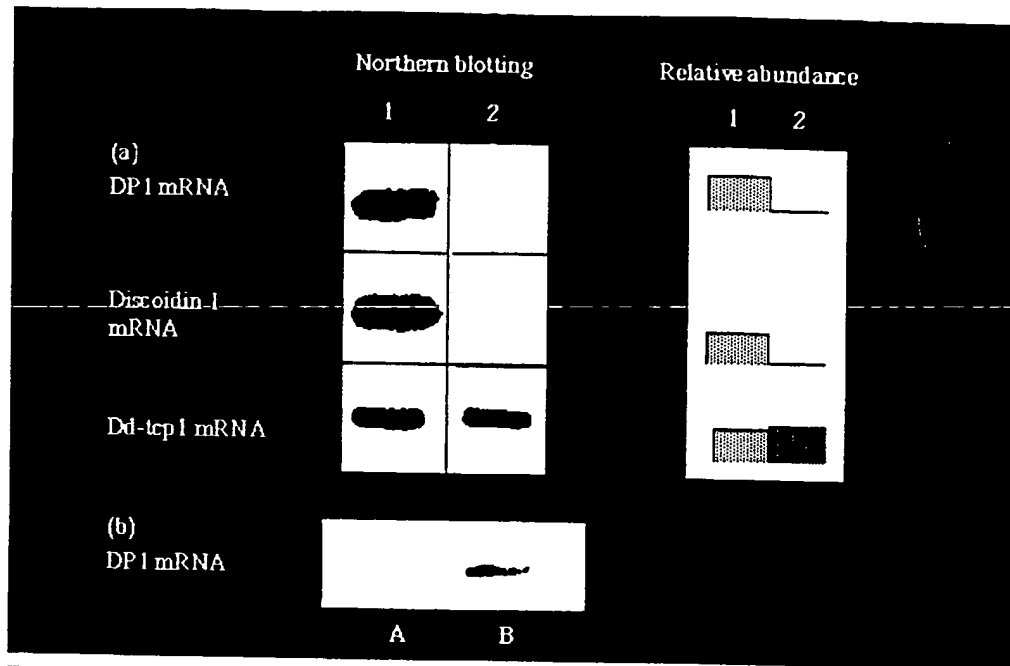
DISC IA NENKVYSLRADYIATALE-  
 DP1 GNNAVYDLTFDYVAVEFNN  
 \* \* \* \* \* \* \* \* \* \*

Sequence ID # 3 deduced amino acid sequence of DP1

Sequence ID # 4 deduced amino acid sequence of Discoidin IA

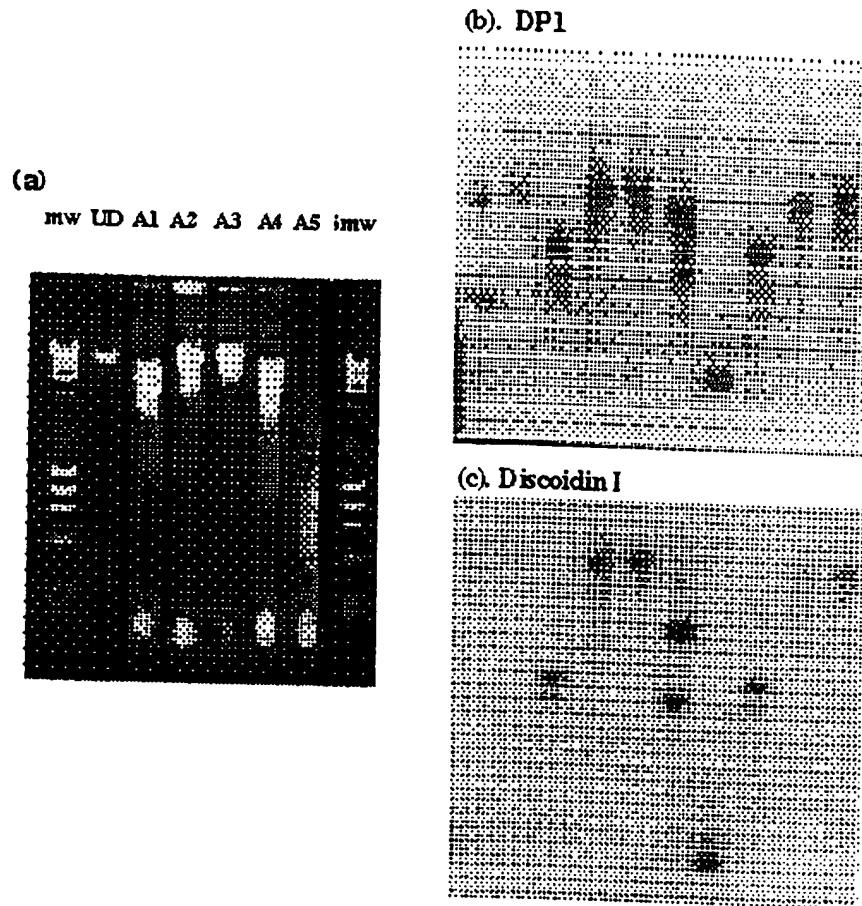
**Figure 4. Pairwise comparison of the deduced amino acid sequences for DP1 and discoidin IA (DISC IA).**

Asterisks (\*) indicate positions of identity while dots (.) indicate positions of conservation. Amino acid sequences obtained for peptides of DP1 are *underlined* and the corresponding numbers are shown in *(parentheses)* underneath the sequences. Regions used for generating primers XJ-1 and XJ-2 are in *italics*. The alignment is performed by use of the multiple alignment program of CLUSTALV (Higgins *et al.*, 1992).



**Figure 5. Northern Blot analyses of DP1 (discoidin II) mRNA expression in axenic strains.**

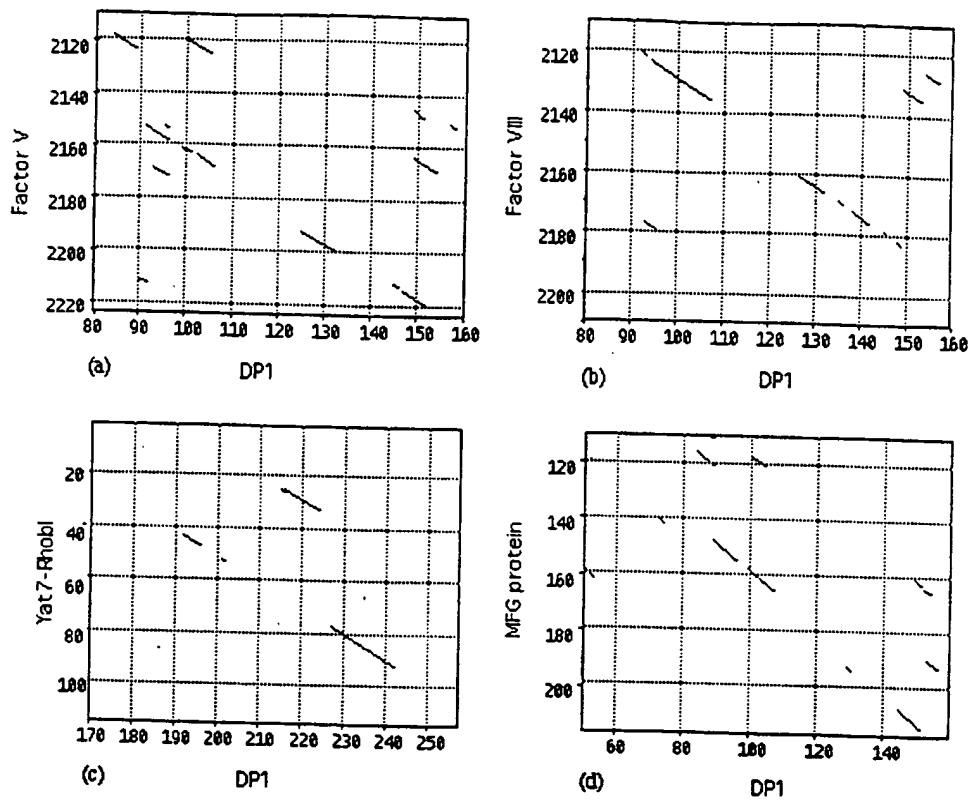
0.5mg samples of mRNA were fractionated on formaldehyde gels. After transfer to Hybond-N membrane, the samples were hybridized to successive hybridizations with  $^{32}\text{P}$ -labeled PCR fragments of DP1 (XJ450), discoidin IA and Dd-tcp1 (control). Lane 1, axenically grown Ax-2; Lane 2, bacterially grown Ax-2. The columns on the right-hand side represent the relative abundance of the hybridized mRNA transcripts in the corresponding lanes of the blots. (b) Hybridization to  $^{32}\text{P}$ -labeled DP1 cDNA XJ-450 of mRNA isolated from amoebae of strain AX-2 harvested from axenic cultures at low density ( $1 \times 10^5$  cells/ml) (A) and at high density ( $4 \times 10^6$  cells/ml) (B).



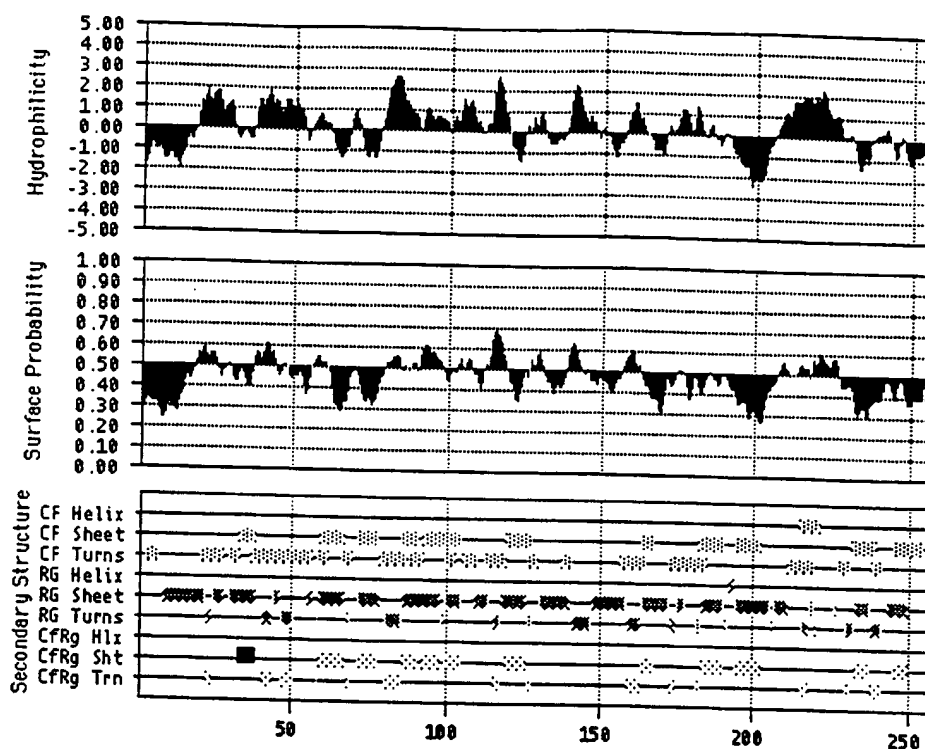
**Figure 6. Southern Blot Analysis of DP1 (discoidin II).** Genomic DNA from the Ax-2 strain is subjected to digestion and the products were separated on a 1% agarose gel (a). After transfer to Hybond-N membrane, the samples were hybridized to the  $^{32}\text{P}$ -labelled XJ-450 probe (b) or the discoidin IA probe (c). Note that in each lane there is only a single band of hybridization to the DP1 (discoidin II) cDNA XJ-450 whereas there were multiple bands for hybridization to the discoidin IA cDNA probe.

Lane identification in (a) is as follows A: Ax-2derived DNA, UD: undigested materials, and mw: Molecular weight marker. Restriction enzymes used in the numbered lanes are: 1. *EcoRI*, 2. *BamHI*, 3. *NotI*, 4. *AccI*, and 5. *Sau3AI*.



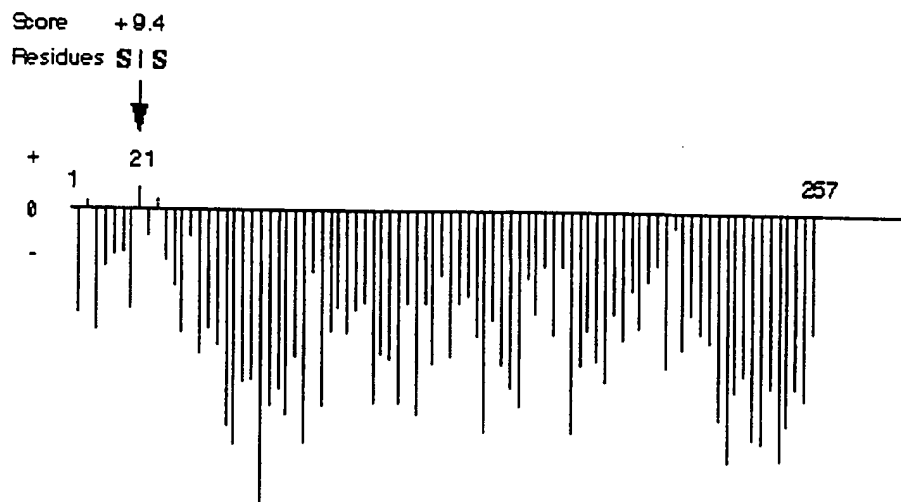


**Figure 7. Sequence Comparison Matrices for DP1 (discoidin II) (horizontal) with (a) human coagulation factor V, (b) human coagulation factor VIII, (c) ORF7 (Yat7-Rhobl) linked to the *Rhodopseudomonas blastica* atp operon and (d) milk fat globule (MFG) protein. The Matrices are plotted using MDM<sub>78</sub> mutation dat matrix Pam250 as the scoring system (Schwartz and Dayhoff 1978), with a window size of 8, and a minimum score of 50%.**

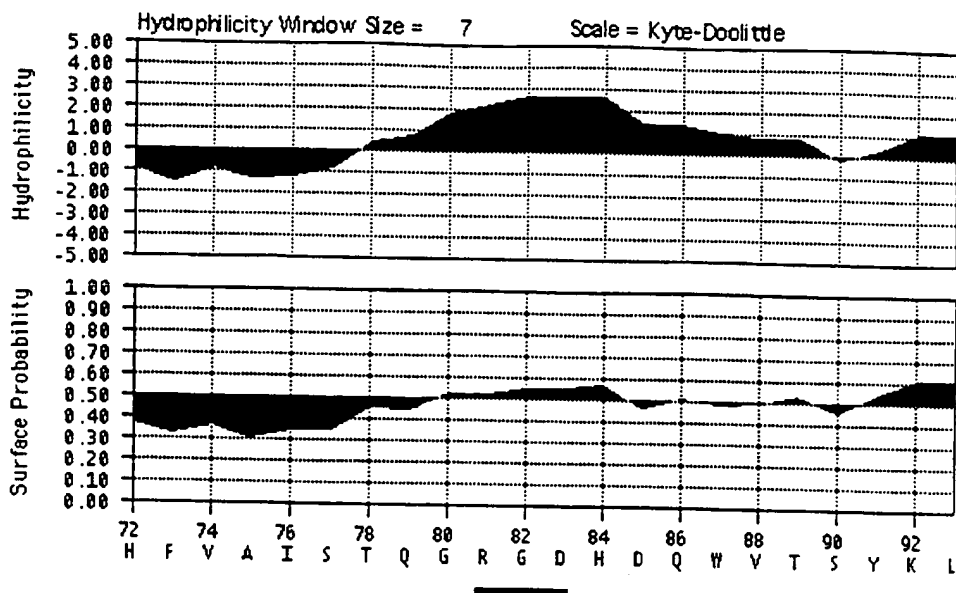


**Figure 8. Hydrophobic plot and the secondary structure prediction for DP1 (discoidin II).**

The profile is constructed using the Kyte-Doolittle algorithm with a sliding window of seven amino acids. Values above the zero axis correspond to hydrophilic segments. Secondary structure predictions based on algorithms of both Chou-Fasman (1978) (CH, *shadowed boxes*) and Robson-Garnier (Garnier *et al.* 1978) (RG, *filled boxes*) are shown in the lower half of the figure.



**Figure 9. Prediction of a probable cleavage site for a putative signal peptide of DP1 (discoidin II) by the method of von Heijne (1986) using the program MacProt (Markiewicz, 1991; Luttko, 1990). A window size of 15 residues of weight matrix for eukaryotic protein is used. Generally a protein having a signal peptide has one segment scoring greater than +3.5 while cytosolic proteins have scores less than +3.5.**



**Figure 10.** Surface probability plot of the RGD-containing region of DP1 (discoidin II).

A surface probability profile of residues 76-88 is constructed using the Janin *et al.* (1978) and Emini *et al.* (1985) algorithms. A probability value above 0.5 is assigned to a water-accessible, "exposed" sequence and values below 0.5 to buried segments.

```

hDP1    NH2-MKVEVLPALTDNYMYLVIDDETKEAAIVDPVQ
          : :                : :: .. :
RSP29-kd NH2-MKiElLPALTDNYMYLiIDedTqxAAvVDPVQ
          ** * ***** ** * ** *****

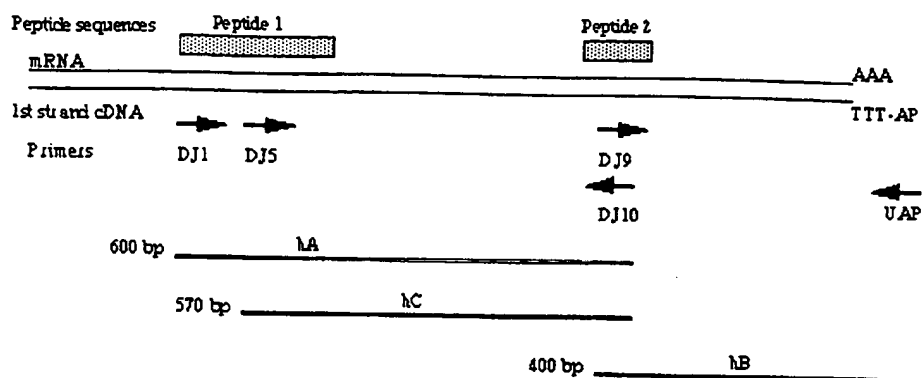
```

Sequence ID# 7 Amino terminal amino acid sequence  
of hDP1

Sequence ID# 8 Amino terminal amino acid sequence  
of rat round spermatid 29,000 Mr protein (RSP-29)

**Figure 11. N-terminal amino acid sequence alignment of hDP1 and the rat round spermatid 29,000 Mr protein (RSP-29).**

A position where both sequences contained the same amino acid is indicated by an *asterisk*. A *colon* indicates a position where the two sequences contain amino acids having similar properties whilst a *dot* indicates a position where the two sequences contain different amino acids.



**Figure 12. Schematic representation for amplification of hDP1 cDNA by the polymerase chain reaction.**

Degenerate primers DJ1, DJ5, DJ9 (sense), and DJ10 (antisense) were designed from a knowledge of peptide sequences of hDP1. The nucleotide sequence of the primer UAP (antisense) had been incorporated into the first strand cDNA during the reverse transcription using primer oligo(dT) 17-AP.

60

ACCGGCCCCGGTTCATGGTGGTGGGCGGAGGGCTGCTCGGCGCGCCGAGCCTCGC -93  
 CGCGCTGGGAGCCGCTGCGCCCGCCGAGGCCTCGGTCCAGCCCTGCTGGGAGTTTTCTGCCA -36  
 CACAGATTTGCGGAAGAACCTGACCGTGGACGAGGGCACCATG AAG GTA GAG GTG CTG 16  
 1 M K V E V L  
 CCT GCC CTG ACC GAC AAC TAC ATG TAC CTG GTC ATT GAT GAT GAG ACC 64  
 7 P A L T D N Y M Y L V I D D E T  
 AAG GAG GCT GCC ATT GTG GAT CCG GTG CAG CCC CAG AAG GTC GTG GAC 112  
 23 V E A A I V D P V Q P Q K V V D  
 GCG GCG AGA AAG CAC GGG GTG AAA CTG ACC ACA GTG CTC ACC ACC CAC 160  
 39 A A R K H G V K L T T V L T T H  
 CAC CAC TGG GAC CAT GCT GGC GGG AAT GAG AAA CTG GTC AAG CTG GAG 208  
 55 H H W D H A G G N E K L V K L E  
 TCG GGA CTG AAG GTG TAC GGG GGT GAC GAC CGT ATC GGG GCC CTG ACT 256  
 71 S G L K V Y G G D D R I G A L T  
 CAC AAG ATC ACT CAC CTG TCC ACA CTG CAG GTG GGG TCT CTG AAC GTC 304  
 87 H K I T H L S T L Q V G S L N V  
 AAG TGC CTG GCG ACC CCG TGC CAC ACT TCA GGA CAC ATT TGT TAC TTC 352  
 103 K C L A T P C H T S G H I C Y F  
 GTG AGC AAG CCC GGA GGC TCG GAG CCC CCT GCC GTG TTC ACA GGT GAC 400  
 119 V S K P G G S E P P A V F T G D  
 ACC TTG TTT GTG GCT GGC TGC GGG AAG TTC TAT GAA GGG ACT GCG GAT 448  
 135 T L F V A G C G K F Y E G T A D  
 GAG ATG TGT AAA GCT CTG CTG GAG GTC TTG GGC CGG CTC CCC CCG GAC 496  
 151 E M C K A L L E V L G R L P P D  
 ACA AGA GTC TAC TGT GGC CAC GAG TAC ACC ATC AAC AAC CTC AAG TTT 544  
 167 T R V Y C G H E Y T I N N L K F  
 GCA CGC CAC GTG GAG CCC GCC AAT GCC GCC ATC CGG GAG AAG CTG GCC 592  
 183 A R H V E P A N A A I R E K L A  
 TGG GCC AAG GAG AAG TAC AGC ATC GGG GAG CCC ACA GTG CCA TCC ACC 640  
 199 W A K E K Y S I G E P T V P S T  
 CTG GCA GAG GAG TTT ACC TAC AAC CCC TTC ATG AGA GTG AGG GAG AAG 688  
 215 L A E E F T Y N P F M R V R E K  
 ACG GTG CAG CAG CAC GCA GGT GAG ACG GAC CCG GTG ACC ACC ATG CGG 736  
 231 T V Q Q H A G E T D P V T T M R  
 GCC GTG CGC AGG GAG AAG GAC CAG TTC AAG ATG CCC CGG GAC TGA GGC 784  
 247 A V R R E K D Q F K M P R D \*  
 CGCCCTGCACCTTCAGCGGATTGGGGATTAGGCTCTTTAGGTAAGTGGCTTTCCTGCTGGT 847  
 CCGTGCGGGAAATTCAGTCTTGATTAAACCTTAATTTTACAGCCCTTGGCTTGTGTTATCGGA 910  
 CGTTTTAATGCATATTTATAAGAGAAGTTTAACAAGTATTTATCCCATAAAAAGGGGGGGC 973  
 CCGTACCCAATTGCGCCTATAGTGAGTCG 1002

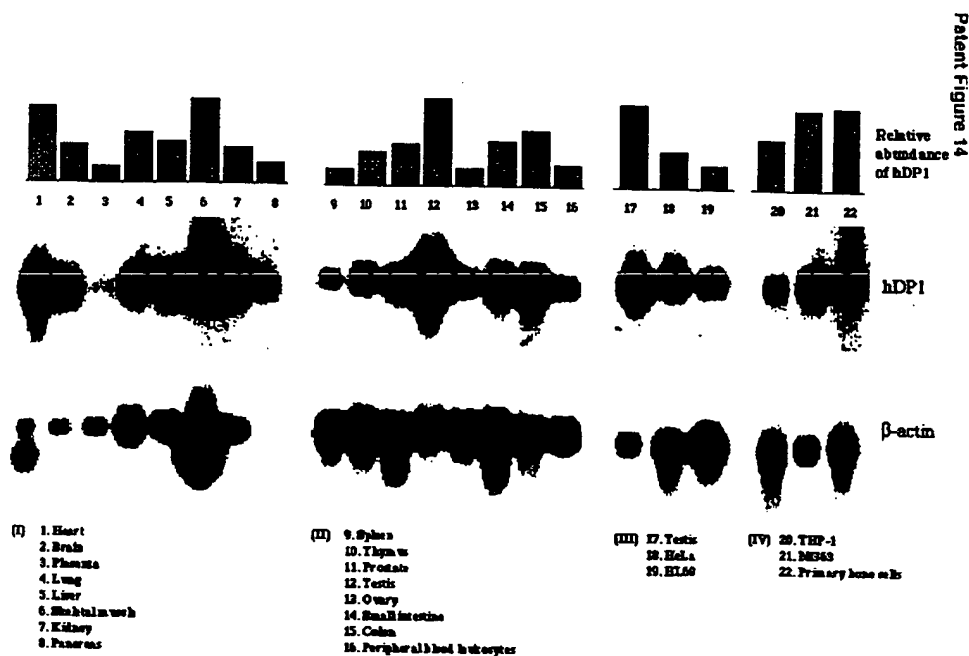
Sequence ID #9 cDNA sequence of HDP1

Sequence ID# 10 deduced amino acid sequence of HDP1

**Figure 13. Nucleotide sequence and deduced amino acid sequence of hDP1 cDNA.**

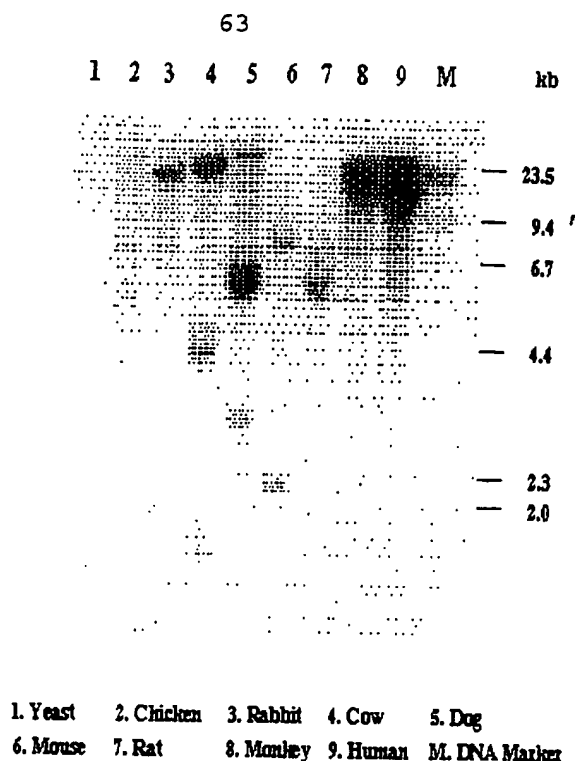
The nucleotide sequence of hDP1 cDNA isolated from a human testis cDNA library consists of 1160 base pairs. The adenine base of the ATG initiation codon is assigned a 1 in the numbering. Nucleotides are numbered in the *right margin* and amino acids on the *left*. An open reading frame of 780-bp encoding 260 amino acids is shown with a *single letter code* for the translated amino acids. A stop codon (TGA) at the terminus of the translation sequence is marked with an *asterisk*. the amino acid sequences of peptides 1 and 2 are *underlined*.





**Figure 14. Northern blot analyses of hDP1 in human tissues.**

Human multiple tissue blots (I and II) counting 2 µg poly(A+)RNA from various tissues were purchased from Clontech. Northern blots III and IV were prepared by running 1 µg poly(A+)RNA in a formaldehyde gel and blotting onto Hybond-N membrane. Hybridization is carried out using a [<sup>32</sup>P]-labeled PCR fragment of hDP1 cDNA. A DNA fragment of β-actin is also labeled with [<sup>32</sup>P] and hybridized to the same blots. The height of the bar on top of each lane represents the relative abundance of hDP1 mRNA in that tissue.

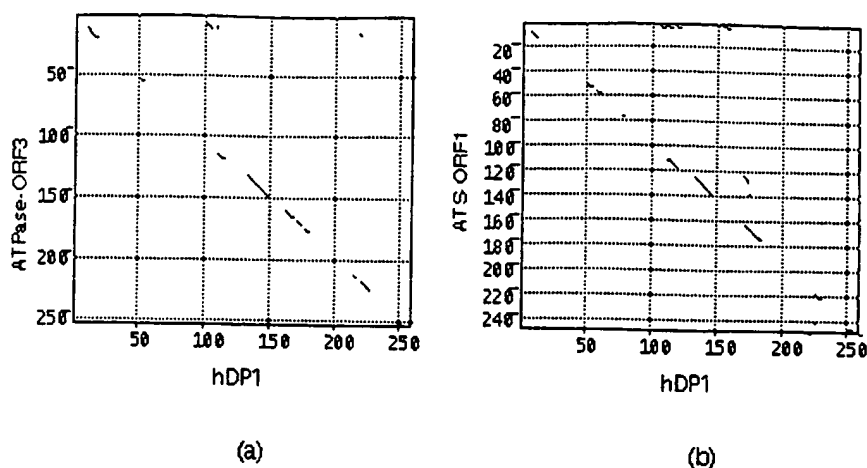


**Fig. 7.13.** Southern analysis of hDP1 on a zoo-blot. The Southern blot containing 8 µg of genomic DNA per lane from nine eucaryotic species was purchased from Clontech. The DNA had been digested with *EcoRI*, run on a 0.7% agarose gel and transferred to a nylon membrane. Hybridization was carried out using a [<sup>32</sup>P]-labelled PCR fragment of hDP1 cDNA.

**Figure 15:** Southern analysis of hDP1 on a zoo-blot. The southern blot containing 8µg og genomic DNA per lane from nine eucaryotic species is obtained from Clontech. The DNA had been digested with *EcoRI*, run on a 0.7% agarose gel and transferred to a nylon membrane. Hybridization is carried out using a [<sup>32</sup>P]-labelled PCR fragment of hDP1 cDNA.



**Figure 16.** Northern Blot analyses of hDP1 homologues in *Dictyostelium discoideum*. A [ $^{32}\text{P}$ ]-labelled PCR fragment of hDP1 cDNA is hybridised to a northern blot containing 5 $\mu\text{g}$  mRNA from *Dictyostelium* amoebae grown on acteria (lane 1), or grown axenically (lane 2).



**Figure 17. Sequence comparison for hDP1.**

Matrix plots of hDP1 (horizontal) against (a). ORF3 linked to the *Rhodopseudomonas blastica atp* operon (ATPase ORF3, vertical) and (b). ORF1 linked to genes for arginyl tRNA synthetase and ribonuclease H of *Buchnera adiphicola* (ATSORF1, vertical). The matrices were plotted using MDM78 mutation data matrix Pam250 as the scoring system (Schwartz and Dayoff, 1978), with a window size of 8, and a minimum score of 55%

Sequence ID #9 cDNA sequence of HDP1

Sequence ID# 10 deduced amino acid sequence of HDP1

```

2 KVEVLPALTDNY.....MYLVID...DETKAAIVDPVQPKVVDAA 40
  : : ::|: :      | | :: | :      ::| :      : ||
33 PDPILGVTEAFKRDTSKKMNLGVGAYRDDNGKSYVLNCVRKAEAMIAA 82

41 RKHGVKLTTLTTHHHWDHAGGNEKLVKLESGLKVYGGDDRIGALTHKIT 90
  :| : |      : :|::: |      ::| :      :| : : |
83 KKMD.KEYLPIAGLADFTASAELALGENSEAFK....SGRYVTV QGIS 126

91 HLSTLQVGSNLNVKCLATPCHTSGHICYFVSKPGGSEPPAVFTGDTLFFVAG 140
  : | || : :      ::      :: ||: : : :| : | : :
127 GTGSLRVGANFLQRFFKFSRD.....VYLPKPSWGNHTPIFRDAGLQLQA 171

141 CGKFYEGTADDEMCKALLEVLGRLPPDTRVY...CGHEYTTNNLKFARHVE 187
  : : |      : :| ::::| :      |:|: | : :      |
172 YRYYDPKTCSLDFTGAMEDISKIPEKSIILLHACAHNPTGVDPRQEQQWKE 221

188 PANAAIREKL.....AWAKEKYSIGEPTVPSTLAEFTY...NPFM 225
  | : |      ::| : : : : : : : : : :
222 LASVVKRNLLAYFDMAYQGFGASGDINRDAWALRHVFIEQGIDVVLSQSYA 271

226 RVREKTQQHAGETDPVTTMRAVRREKQDFKM 257
  : : : : : | : | | :| :| :
272 KNMGLYGERAGAFTVICRDAEEAKRVESQLKI 303

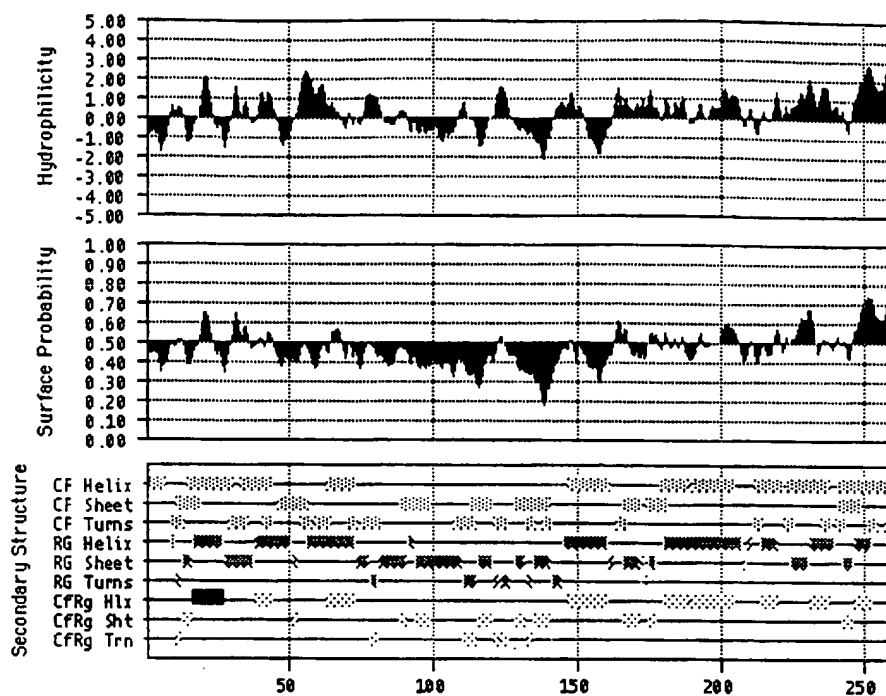
```

sequence ID#9 amino acid sequence of HDP1

sequence ID#11 amino acid sequence of aspartate amino transferase

**Figure 18. Sequence comparison of hDP1 with aspartate aminotransferase (AAT) from chicken mitochondria.**

hDP1 is shown in the upper line and DP1 in the lower one. "|" indicates identity between aligned residues; ":" indicates similarity. The comparison is performed using the program "bestfit" of the GCG package (Devereux *et al.*, 1984). Identity between the two proteins is 14.3% and the similarity is 44.1%.



**Figure 19. Hydrophilic plot and secondary structure prediction for hDP1.**

The hydrophilic profile is constructed using the Kyte-Doolittle algorithm (Kyte and Doolittle, 1982) with a sliding window of seven amino acids. *Values above the zero axis correspond to hydrophilic segments.* Secondary structure prediction based on the algorithms of Chou-Fasman (1978) (CH, *shadowed boxes*) and Robinson-Garnier (Garnier *et al.* 1978) (RG, *filled boxes*) are shown in the lower half of the figure.

```

16 LVIDDETKAAIVDPVQPQKVVDAAARKHGVKLTTLTTHHHWDHAGGNEK 65
   : : : : : | : : : : :
1 MSVPAGSVSCLANALLNLRSSSTDYNADHGVKNSIL.....NFSNSKDASR 45

66 LVKLES.GLKVYGGDDRIGALTHKITHLSTLQ.....VGSLNVKC 104
   : | : | : : : | : : : : | : : :
46 FDGSESWSSSVLDKNQFIVAGSDSVKHFVAISTQGRGDHDQWVTSYKLRY 95

105 .LATPCHTSGHICYFVSKPGGSEPPAVFTGDTLFFVAGCGKFYEGTADEMC 153
    | : : : : : : : : : : : :
96 TLDNVNWVEYNNGEIIINANKDRNSIVTINFNPPIKARSIAIHPQTYNNGHI 145

154 KALLEVLGRLPPDTRVYCGHEYTINNLKFARHVEPANAAIRE...KLAWA 200
   : | : | : | : : : : : : : : :
146 SLRWELYA.LP..VKSYSNPSVQVGEVSIGDRSLNSGTGSRTIVRHVKFP 192

201 KEKYSIGEPTVPSTLAEFTYNPFMRVREKTVQQHAGETDPVTMRAVRR 250
    | | : : : | | | : : : : :
193 VEFLSVPIVSIGCKKVAHTDNGQMR.WEGKSENITTKGFDLTFI..TWG 239

251 EKDQFKMPRD 260
   : : : |
240 NNAVYDLTFD 249

```

Sequence ID# 10 amino acid sequence of HDP1

Sequence ID # 2 amino acid sequence of discoidin II

**Figure 20. Amino acid sequence comparison of hDP1 with DP1 (discoidin II).**

hDP1 is shown in the upper line and DP1 in the lower one. "|" indicates identity between aligned residues; ":" indicates similarity. The comparison is performed using the program "bestfit" of the GCG package (Devereux *et al.*, 1984). The identity between the two proteins is 12.8% and the similarity is 38.5%.

CATAATGAAAGTTATTTTCGTAGTTTGTAGCCATTGTATTAGTTACATTATGGGCT

56 ATGCCATCAGAAGCTGGTAAAGACCCAAAGATTACCAATAAAGTATTCTTTGATATAGAA  
M P S E A G K D P K I T N K V F F D I E 20

116 ATTGATAATAAACCAGCAGGTAGAATTGTATTTGGTTTATATGGAAAGACAGTACCAAAA  
I D N K P A G R I V F G L Y G K T V P K 40

176 ACAGTTGAAAACTTTAGAGCATTATGTACTGGTGAAAAAGGTTTAGGTACCAGTGGTAAA  
T V E N F R A L C T G E K G L G T S G K 60

236 CCATTACATTATAAAGATAGTAAATTCATCGTATCATTCCAAACTTTATGATTCAAGGT  
P L H Y K D S K F H R I I P N F M I Q G 80

296 GGTGATTTACAAGAGGTGATGGTACTGGTGGTGAATCAATTTATGGTAAAAAATTCAAT  
G D F T R G D G T G G E S I Y G K K F N 100

356 GATGAAAACCTTCAAATTAACACTCCAAACCAGGTCTTTTATCAATGGCTAACGCTGGT  
D E N F K I K H S K P G L L S M A N A G 120

416 CCAAACACTAATGGTTCACAATCTTTTATTACTACCGTTGTTACTTCATGGTTAGATGGT  
P N T N G S Q F F I T T V V T S W L D G 140

476 CGTCATACTGTTTTTGGTGAAGTTATTGAAGGTATGGATATTGTTAAACTCCTTGAATCC  
R H T V F G E V I E G M D I V K L L E S 160

536 ATTGGTTCCCAATCTGGAACACCAAGTAAATTGCTAAATCTCAAACCTCTGGTGAATTA  
I G S Q S G T P S K I A K I S N S G E L 180

596 TAAATAAAATAAAACCAAACCAAATAAAATAAAT

Sequence ID#12 nucleotide cDNA sequence of DdCyP2

Sequence ID#13 predicted amino acid sequence of DdCyP2

**Figure 21. Nucleotide sequence of the Wt.3 cDNA insert and the deduced amino acid sequence of Dd CyP2.**

The nucleotide sequence shown is for the 629 bp cDNA of Dd CyP2 in clone Wt.3. Nucleotides are numbered on the left and amino acids on the right. An open reading frame of 540 bp encoding 180 amino acids is shown using the single letter code for the translated amino acids. The initiation codon (ATG) and the termination codon (TAA) are underlined. An RGD motif is shown in bold-type. The five N-terminal amino acid residues absent from the isolated Dd CyP2 protein are in italics.



1					50
1	MPSEAGKDPK	ITNKVFFDIE	IDNKPAGRIV	FGLYGKTVPK	TVENFRALCT
2	.MTTVKPTSP	ENPRVFFDIT	IGGVEAGKV	MELYANTVPK	TAENFRALCT
3	.....M	VNPKVFFDMT	VGDKAAGRIV	MELYADTVPE	TAENFRALCT
4	.....	..SQVYFDVE	ADGQPIGRV	FKLYNDIVPK	TAENFRALCT
5	.....	VNPTVFFDIA	VDGEPLGRVS	FELFADKVPK	TAENFRALST
51					100
1	GEKGLGTSGK	PLHYKDSKFH	RIITNFMIOG	GDFTRGDGTG	GESIYGKKFN
2	GEKGIGKSGK	PLSYKGSSFH	RVITNFMCOG	GDFTMGNGTG	GESIYGKFKFA
3	GERGIGKSGK	PLHYKGSAPH	RVIPKFMCOG	GDFTAGNGTG	GESIYGMKFK
4	GEKGFG.....	..YAGSPFH	RVIPDFMLQG	GDFTAGNGTG	GKSIYGGKFP
5	GEKGFG.....	..YKGSCFH	RIIPGFMCQG	GDFTRHNGTG	GKSIYGEKFE
			* * * *		
			+ + ++ +	++	
101					150
1	DENFKIKHSK	PGLLSMRNAG	PNTNGSQFFI	TTVVTSWLDG	RHTVFGFVIE
2	DENFKLKHFG	QGTLSMANAG	ANTNGSQFFI	CVAPTDWLDG	KHVVFQFVTE
3	DENFVKKHTG	PGILSMRAG	SNTNGSQFFI	CTEKTSWLDG	KHVVFQGVVE
4	DENFKKHHR	PGLLSMANAG	PNTNGSQFFI	TTVPCPWLDG	KHVVFGEVVD
5	DENFILKHTG	PGILSMANAG	PNTNGSQFFI	CTAKTEWLDG	KHVVFQKVKE
		**	* *	*	*
		+++	+ +	++	+
151			180	Identity with Dd CyP2	
1	GMDIVKLLES	IGSQSGTPSK	IAKISNSGEL		
2	GMDVVKMEA	AGSQSGKTK	PVVIANCGQL	65.0%	in 175 aa overlap
3	GMDVVRDIEK	VGSDSGRTSK	KVVTCDGQL	65.9%	in 170 aa overlap
4	GYDIVKKVES	LGSPSGATKA	RIVVAKSGEL	64.9%	in 168 aa overlap
5	GMNIVEAMER	FGSRNGKTSK	KITIADCGQLE	61.8%	in 170 aa overlap
		*			

**Figure 22. Alignment of the amino acid sequence of Dd CyP2 and the sequences of selected members of the cyclophilin A family of proteins.**

The sequences were extracted from the updated releases from GenBank and Swissprot. The alignment is performed using "pileup" of the GCG program (the Wisconsin Genetic Computer Group). The sequences of the CyPs were derived from the following sources: (1) Dd CyP2; (2) Dd CyP1 (Barisic *et al.*, 1991); (3) *Brassica napus* (p24525, Gasser *et al.*, 1990); (4) *Saccharomyces cerevisiae* (p14832, Haendler *et al.*, 1989); (5) human (p05092, Haendler *et al.*, 1987). Residues of human CyPA that are located in close contact with a tetrapeptide Ala-Ala-Pro-Ala substrate are marked with an asterisk (Kallen & Walkinshaw, 1992); residues of human CyPA that are in close contact with bound cyclosporin A are marked with a cross (Therriault *et al.*, 1993; Pflugl *et al.*, 1993).

1	50
1	..... ..MVNPTVFF
2	MLALRCGSRW LGLLSVPRSV PLRLPAARAC SKGSGDPSSS SSSGNPLVYL
3	..... ..MPSEAGK DPKITNKVFF
4	..... MKVLLAAALI AGSVFFLLLP GPSAADEKKK GPKVTVKVYF
5	.....MG PGPRLLLPLV LCVGLGALVF SSGAEGFRKR GPSVTAKVFF
51	100
1	DIAVDGEPLG RVSFELFADK VPKTAENFRA LSTGEKGF... ..GYKGS
2	DVDANGKPLG RVVLELKADV VPKTAENFRA LCTGEKGF... ..GYKGS
3	DIEIDNKPAG RIVFGLYGKT VPKTVENFRA LCTGEKGLGT SGKPLHYKDS
4	DLRIGDEDVG RVIFGLFGKT VPKTVDNFVA LATGEKGF... ..GYKNS
5	DVRIGDKDVG RIVIGLFGKV VPKTVENFVA LATGEKGY... ..GYKGS
101	150
1	CFHRIIPGFM CQGGDFTRHN GTGGKSIYGE KFEDENFILK HTGPGILSMA
2	TFHRVIPFSM CQAGDFTNHN GTGGKSIYGS RFPDENFTLK HVGPGVLSMA
3	KFHRIITNFM IQGGDFTRGD GTGGESIYGK KFNDENFKIK HSKPGLLSMR
4	KFHRVIKDFM IQGGDFTRGD GTGGKSIYGE RFPDENFKLK HYGPGWVSMA
5	KFHRVIKDFM IQGGDITTGD GTGGVSIYGE TFPDENFKLK HYGIGWVSMA
151	200
1	NAGPNTNGSQ FFICTAKTEW LDGKHVVFGK VKEGMNIVEA MERFGSRNGK
2	NAGPNTNGSQ FFICTIKTDW LDGKHVVFGH VKEGMDVVKK IESFGSKSGR
3	NAGPNTNGSQ FFITTVVTSW LDGKHVVFGK VIEGMDIVKL LESIGSQSGT
4	NAGKDTNGSQ FFITTVKTAW LDGKHVVFGK VLEGMEVVRK VESTKTDSRD
5	NAGPDTNGSQ FFITLTKPTW LDGKHVVFGK VIDGMTVVHS IELQATDGH
201	Identity with Dd CyP2 (3)
1	TSKKITIADC GQLE..... 61.8% in 170 aa
2	TSKKIVITDC GQLS..... 61.4% in 166 aa
3	PSKIAKISNS GEL ..... 60.6% in 180.aa
4	KPLKDVIIAD CGKIEVEKPF AIAKE.. 60.7% in 166 aa
5	RPLTNCIIIN SGKIDVKTPF VVEIADW

**Figure 23. Alignment of the amino acid sequences of Dd CyP2 and four human cyclophilins.**

The sequences were extracted from the updated releases from GenBank and Swissprot. The alignment is performed using "pileup" of the GCG program (the Wisconsin Genetic Computer Group). The sequences of the CyPs were derived from the following sources: (1) human CyPA (p05092, Haendler *et al.*, 1987), (2) human CyPD (p30405, Bergsma *et al.*, 1991); (3) Dd CyP2; (4) human CyPB (p23284, Price *et al.*, 1991); (5) human CyPC (Schneider *et al.*, 1994).

50

100

100

150

150

200

200

### Identity with Dd Cyp2

60.6% in 180 aa

61.9% in 181 aa

61.1% in 180 aa

54.9% in 182 aa

102 55

The sequences were extracted from the updated releases from GenBank and Swissprot. The alignment is performed using "pileup" of the GCG program (the Wisconsin Genetic Computer Group). The sequences of the CyPBs were derived from the following sources: (1) human (p23284, Price *et al.*, 1991); (2) mouse (p24369, Hasel *et al.*, 1991); (3) chick (p24367, Caroni *et al.*, 1991); (4) rat (p24368, Iwai and Inagami *et al.*, 1990). (5) Dd CyP2 (XP1). Potential hydrophobic signal sequences of CyPBs are underlined. The RGD motifs are in bold-type.

```

1
1 ..... .MAFPKVYFD MTIDGQPAGR IVMElyTDKT PRtAENFRAL
2 ..... .MANPKVFFD LTIGGAPAGR VVMELFADTT PKtAENFRAL
3 ..... ....PKVYFD MTVGDKAAGR IVMElyADTV PETaENFRAL
4 ..... ..... ..MELFQDVV PQtAENFRAL
5 ..... .MANPRVFFD MTVGGAPAGR IVMElyANEV PKtAENFRAL
6 ..MPSEAGKD PKITNKVFFD IEIDNKPAGR IVFGLYGKTV PKtVENFRAL
7 ..... ...MAHCFFD MTIGGQPAGR IMELFPD.V PKtAENFRAL
8 AEEEEVIEPQ AKVTNKVYFD VEIGGEVAGR IVMGLFGEVV PKtVENFRAL
9 ...MTTVKPT SPENPRVFFD ITIGGVEAGK VVMELYANTV PKtAENFRAL

```

51

```

1 CTGEKGVGGT GKPLHFKGSK FHRVIPNFM C QGGDFTAGNG TGGESIYGSK
2 CTGEKGVGKM GKPLHYKGST FHRVIPGFMC QGGDFTAGNG TGGESIYGAK
3 CTGERGIGKS GKPLHYKGSA FHRVIPKFMC QGGDFTAGNG TGGESIYGMK
4 CTGEKGMGDR .KPLHYKGSS FHRVIPGFMC QGGDFTAGNG TGGESIYGAK
5 CTGEKGVGKS GKPLHYKGST FHRVIEPFMC QGGDFTAGNG TGGESIYGEK
6 CTGEKGLGTS GKPLHYKDSK FHRIIPNFM QGGDFTAGNG TGGESIYGGK
7 CTGEKGIGPS GKMTYEGSV FHRVIPKFML QGGDFTAGNG TGGESIYGAK
8 CTGEKKYG.. ....YKGSS FHRIIKDFM QGGDFTAGNG TGGISIYGAK
9 CTGEKGIGKS GKPLSYKGSS FHRVITNFM C QGGDFTAGNG TGGESIYGNK

```

101

```

1 FEDENFERKH TGPGLSMAN AGANTNGSQF FICTVKTDWL DGKHVVFGQV
2 FNDENFVKKH TGPGLSMAN AGPGTNGSQF FICTAKTEWL NGKHVVFGQV
3 FKDENFVKKH TGPGLSMRN AGSNTNGSQF FICTEKTSWL DGKHVVFGQV
4 FKDENFIKKH TGPGLSMAN AGPGTNGSQF FICTEKTAWL DGKHVVFGQV
5 FPDEKFVRKQ PAPGVLSMAN AGPNTNGSQF FICTVATPWL DGKHVVFGQV
6 FNDENFKIKH SKPGLLSMAN AGPNTNGSQF FITTVVTSWL DGRHTVFGEV
7 FADENFIKKH TTPGLLSMAN AGPGTNGSQF FITTVATPHL DGKHVVFGKV
8 FEDENFTLKH TGPGLSMAN AGPNTNGSQF FICTVKTSWL DNKHVVFGQV
9 FADENFKLKH FGQGLSMAN AGANTNGSQF FICVAPTDWL DGKHVVFGFV

```

151

Identity with Dd CyP2

1	VEGLDVVKAI	EKVGSSSG.K	PTKPVVADC	GQLS	67.7% in 167 aa
2	VEGMDVIKKA	EAVGSSSG.R	CSKPVVIADC	GQL	67.1% in 167 aa
3	VEGMDVVRDI	EKVGSDSG.R	TSKKVVT CDC	GQL	65.9% in 170 aa
4	VEGMDVVRAI	EKVGSGSG.Q	TKKPVKIADC	GQLS	67.6% in 148 aa
5	VEGMDVVKAI	EKVGTNRNG.S	TSKVVKVADC	GQLS	67.7% in 167 aa
6	IEGMDIVKLL	ESIGSQSG.T	PSKIAKISNS	GEL	
7	VEGMDVVRKI	EATQTDGRGD	PLSEVKIAKC	GQL	63.3% in 166 aa
8	IEGMDVVRKI	ESQETRAFDV	PKKGCRIVAC	GELPLDA	64.9% in 174 aa
9	TEGMDVVKKM	EAAGSQSGKT	TKPVVIANCG	QL	65.0% in 175 aa

**Figure 25.** Alignment of the amino acid sequences of two Dd CyPs and some plant CyPs.

The sequences were extracted from the updated releases from GenBank and Swissprot. The alignment is performed using "pileup" of the GCG program (the Wisconsin Genetic Computer Group). The sequences for the CyPs were derived

from the following sources: (1) *Arabidopsis thaliana* (114844, Lippuner *et al.*, 1994) (2) tomato (m55019, Gasser *et al.*, 1990); (3) *Brassica napus* (m55018, Gasser *et al.*, 1990); (4) onion (113365, Clark *et al.*, 1993); (5) maize (m55021, Gasser *et al.*, 1990); (6) Dd CyP2 (XP1); (7) *Arabidopsis thaliana* (x63616, Bartling *et al.*, 1991); (8) *Arabidopsis thaliana* (114845, nuclear-encoded chloroplast stromal, Lippuner *et al.*, 1994); (9) Dd CyP1 (Barisic *et al.*, 1991). The seven amino acid insertion in Dd CyP2 is in bold-type. The potential ATP/GTP binding sites are underlined.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: COOK, JONATHAN S  
EBETINO, FRANK H  
IBBOTSON, KENNETH J  
JI, XIAOHUI  
ROGERS, MICHAEL J  
RUSSELL, ROBERT G  
XIONG, XIAOJUAN
- (ii) TITLE OF INVENTION: BISPHOSPHONATE BINDING PROTEINS
- (iii) NUMBER OF SEQUENCES: 34
- (iv) CORRESPONDENCE ADDRESS:  
(A) ADDRESSEE: THE PROCTER & GAMBLE COMPANY  
(B) STREET: 11801 EAST MIAMI RIVER ROAD  
(C) CITY: ROSS  
(D) STATE: OHIO  
(E) COUNTRY: USA  
(F) ZIP: 45061
- (v) COMPUTER READABLE FORM:  
(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:  
(A) APPLICATION NUMBER:  
(B) FILING DATE:  
(C) CLASSIFICATION:
- (vii) ATTORNEY/AGENT INFORMATION:  
(A) NAME: MAKE, RICHARD A  
(B) REGISTRATION NUMBER: 37,343
- (ix) TELECOMMUNICATION INFORMATION:  
(A) TELEPHONE: (513) 627-0087  
(B) TELEFAX: (513) 627-0260

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 968 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (ix) FEATURE:  
(A) NAME/KEY: CDS  
(B) LOCATION: 61..831

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

TTTTATTAGT ATTTAATTTA TTTATATAAT TCTTTTAA AAAAAACAAA ACAAACAAA      60
ATG TCC GTT CCA GCT GGT TCT GTT TCA TGT CTT GCT AAT GCA TTA TTA      108
Met Ser Val Pro Ala Gly Ser Val Ser Cys Leu Ala Asn Ala Leu Leu
  1           5           10           15

AAT TTA AGA TCA TCA ACT GAT TAT AAT GCT GAT CAT GGT GTA AAG AAT      156
Asn Leu Arg Ser Ser Thr Asp Tyr Asn Ala Asp His Gly Val Lys Asn
  20           25           30

TCT ATT TTA AAT TTT TCA AAT TCA AAG GAT GCT AGT AGA TTC GAC GGT      204

```

76

Ser Ile Leu Asn Phe Ser Asn Ser Lys Asp Ala Ser Arg Phe Asp Gly	
35 40 45	
AGT GAA TCA TGG TCA TCA TCA GTT TTG GAT AAG AAT CAA TTC ATT GTT	252
Ser Glu Ser Trp Ser Ser Ser Val Leu Asp Lys Asn Gln Phe Ile Val	
50 55 60	
GCC GGT AGT GAT TCT GTT AAA CAT TTC GTT GCA ATC TCA ACT CAA GGT	300
Ala Gly Ser Asp Ser Val Lys His Phe Val Ala Ile Ser Thr Gln Gly	
65 70 75 80	
CGT GGT GAT CAT GAT CAA TGG GTA ACT TCA TAC AAA TTA AGA TAC ACA	348
Arg Gly Asp His Asp Gln Trp Val Thr Ser Tyr Lys Leu Arg Tyr Thr	
85 90 95	
CTT GAT AAT GTA AAC TGG GTT GAA TAT AAC AAT GGT GAA ATA ATC AAT	396
Leu Asp Asn Val Asn Trp Val Glu Tyr Asn Asn Gly Glu Ile Ile Asn	
100 105 110	
GCC AAT AAA GAT AGA AAT TCA ATT GTT ACA ATC AAC TTT AAT CCA CCA	444
Ala Asn Lys Asp Arg Asn Ser Ile Val Thr Ile Asn Phe Asn Pro Pro	
115 120 125	
ATT AAA GCT AGA TCT ATT GCC ATT CAT CCT CAA ACC TAT AAT AAT CAT	492
Ile Lys Ala Arg Ser Ile Ala Ile His Pro Gln Thr Tyr Asn Asn His	
130 135 140	
ATT TCA CTT CGT TGG GAA TTA TAT GCA TTA CCA GTT AAA AGT TAT TCA	540
Ile Ser Leu Arg Trp Glu Leu Tyr Ala Leu Pro Val Lys Ser Tyr Ser	
145 150 155 160	
AAT CCA TCA GTC CAA GTT GGT GAA GTT TCA ATT GGT GAT AGA TCT CTT	588
Asn Pro Ser Val Gln Val Gly Glu Val Ser Ile Gly Asp Arg Ser Leu	
165 170 175	
AAC AGT GGT ACT GGT TCA CGT ACG ATT GTT CGT CAC GTT AAA TTC CCA	636
Asn Ser Gly Thr Gly Ser Arg Thr Ile Val Arg His Val Lys Phe Pro	
180 185 190	
GTG GAA TTC CTT TCT GTT CCA ATC GTA TCA ATT GGT TGT AAA AAA GTT	684
Val Glu Phe Leu Ser Val Pro Ile Val Ser Ile Gly Cys Lys Lys Val	
195 200 205	
GAT GCA CAT ACT GAT AAT GGT CAA ATG AGA TGG GAA GGT AAA TCT GAA	732
Asp Ala His Thr Asp Asn Gly Gln Met Arg Trp Glu Gly Lys Ser Glu	
210 215 220	
AAT ATT ACT ACA AAA GGT TTT GAT TTA ACT TTT ATT ACA TGG GGT AAT	780
Asn Ile Thr Thr Lys Gly Phe Asp Leu Thr Phe Ile Thr Trp Gly Asn	
225 230 235 240	
AAT GCA GTT TAT GAT TTA ACT TTT GAT TAT GTT GGT GTT GAA TTT AAT	828
Asn Ala Val Tyr Asp Leu Thr Phe Asp Tyr Val Ala Val Glu Phe Asn	
245 250 255	
AAT TAAATAATTA AATAATAAAA TAAATAAATA AATTTATTIG TTTTATTTT	881
Asn	
ATATTTTAAA ATAATTAAT AATTAATAAA TTAAAAAAAA AAAAAAAAAA AAAATTTTAA	941
AATTTTCCAG AAAAAAAAAA AAAAAAA	968

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 257 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

77

Met Ser Val Pro Ala Gly Ser Val Ser Cys Leu Ala Asn Ala Leu Leu  
 1 5 10 15  
 Asn Leu Arg Ser Ser Thr Asp Tyr Asn Ala Asp His Gly Val Lys Asn  
 20 25 30  
 Ser Ile Leu Asn Phe Ser Asn Ser Lys Asp Ala Ser Arg Phe Asp Gly  
 35 40 45  
 Ser Glu Ser Trp Ser Ser Ser Val Leu Asp Lys Asn Gln Phe Ile Val  
 50 55 60  
 Ala Gly Ser Asp Ser Val Lys His Phe Val Ala Ile Ser Thr Gln Gly  
 65 70 75 80  
 Arg Gly Asp His Asp Gln Trp Val Thr Ser Tyr Lys Leu Arg Tyr Thr  
 85 90 95  
 Leu Asp Asn Val Asn Trp Val Glu Tyr Asn Asn Gly Glu Ile Ile Asn  
 100 105 110  
 Ala Asn Lys Asp Arg Asn Ser Ile Val Thr Ile Asn Phe Asn Pro Pro  
 115 120 125  
 Ile Lys Ala Arg Ser Ile Ala Ile His Pro Gln Thr Tyr Asn Asn His  
 130 135 140  
 Ile Ser Leu Arg Trp Glu Leu Tyr Ala Leu Pro Val Lys Ser Tyr Ser  
 145 150 155 160  
 Asn Pro Ser Val Gln Val Gly Glu Val Ser Ile Gly Asp Arg Ser Leu  
 165 170 175  
 Asn Ser Gly Thr Gly Ser Arg Thr Ile Val Arg His Val Lys Phe Pro  
 180 185 190  
 Val Glu Phe Leu Ser Val Pro Ile Val Ser Ile Gly Cys Lys Lys Val  
 195 200 205  
 Asp Ala His Thr Asp Asn Gly Gln Met Arg Trp Glu Gly Lys Ser Glu  
 210 215 220  
 Asn Ile Thr Thr Lys Gly Phe Asp Leu Thr Phe Ile Thr Trp Gly Asn  
 225 230 235 240  
 Asn Ala Val Tyr Asp Leu Thr Phe Asp Tyr Val Ala Val Glu Phe Asn  
 245 250 255  
 Asn

## (2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 317 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Gln Phe Ile Val Ala Gly Ser Asp Ser Val Lys His Phe Val Ala Ile  
 1 5 10 15  
 Ser Thr Gln Gly Arg Gly Asp His Asp Gln Trp Val Thr Ser Tyr Lys  
 20 25 30  
 Leu Arg Tyr Thr Leu Asp Asn Val Asn Trp Val Glu Tyr Asn Asn Gly  
 35 40 45



78

Glu Ile Ile Asn Ala Asn Lys Asp Arg Asn Ser Ile Met Ser Val Pro  
 50 55 60  
 Ala Gly Ser Val Ser Cys Leu Ala Asn Ala Leu Leu Asn Leu Arg Ser  
 65 70 75 80  
 Ser Thr Asp Tyr Asn Ala Asp His Gly Val Lys Asn Ser Ile Leu Asn  
 85 90 95  
 Phe Ser Asn Ser Lys Asp Ala Ser Arg Phe Asp Gly Ser Glu Ser Trp  
 100 105 110  
 Ser Ser Ser Val Leu Asp Lys Asn Gln Phe Ile Val Ala Gly Ser Asp  
 115 120 125  
 Ser Val Lys His Phe Val Ala Ile Ser Thr Gln Gly Arg Gly Asp His  
 130 135 140  
 Asp Gln Trp Val Thr Ser Tyr Lys Leu Arg Tyr Thr Leu Asp Asn Val  
 145 150 155 160  
 Asn Trp Val Glu Tyr Asn Asn Gly Glu Ile Ile Asn Ala Asn Lys Asp  
 165 170 175  
 Arg Asn Ser Ile Val Thr Ile Asn Phe Asn Pro Pro Ile Lys Ala Arg  
 180 185 190  
 Ser Ile Ala Ile His Pro Gln Thr Tyr Asn Asn His Ile Ser Leu Arg  
 195 200 205  
 Trp Glu Leu Tyr Ala Leu Pro Val Lys Ser Tyr Ser Asn Pro Ser Val  
 210 215 220  
 Gln Val Gly Glu Val Ser Ile Gly Asp Arg Ser Leu Asn Ser Gly Thr  
 225 230 235 240  
 Gly Ser Arg Thr Ile Val Arg His Val Lys Phe Pro Val Glu Phe Leu  
 245 250 255  
 Ser Val Pro Ile Val Ser Ile Gly Cys Lys Lys Val Asp Ala His Thr  
 260 265 270  
 Asp Asn Gly Gln Met Arg Trp Glu Gly Lys Ser Glu Asn Ile Thr Thr  
 275 280 285  
 Lys Gly Phe Asp Leu Thr Phe Ile Thr Trp Gly Asn Asn Ala Val Tyr  
 290 295 300  
 Asp Leu Thr Phe Asp Tyr Val Ala Val Glu Phe Asn Asn  
 305 310 315

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 313 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Gln Tyr Ile Val Ala Gly Cys Glu Val Pro Arg Thr Phe Met Cys Val  
 1 5 10 15  
 Ala Leu Gln Gly Arg Gly Asp Ala Asp Gln Trp Val Thr Ser Tyr Lys  
 20 25 30  
 Ile Arg Tyr Ser Leu Asp Asn Val Ser Trp Phe Glu Tyr Arg Asn Gly  
 35 40 45

79

Ala Ala Val Thr Gly Val Thr Asp Arg Asn Thr Val Met Ser Thr Gln  
 50 55 60  
 Gly Leu Val Gln Leu Leu Ala Asn Ala Gln Cys His Leu Arg Thr Ser  
 65 70 75 80  
 Thr Asn Tyr Asn Gly Val His Thr Gln Phe Asn Ser Ala Leu Asn Tyr  
 85 90 95  
 Lys Asn Asn Gly Thr Asn Thr Ile Asp Gly Ser Glu Ala Trp Cys Ser  
 100 105 110  
 Ser Ile Val Asp Thr Asn Gln Tyr Ile Val Ala Gly Cys Glu Val Pro  
 115 120 125  
 Arg Thr Phe Met Cys Val Ala Leu Gln Gly Arg Gly Asp Ala Asp Gln  
 130 135 140  
 Trp Val Thr Ser Tyr Lys Ile Arg Tyr Ser Leu Asp Asn Val Ser Trp  
 145 150 155 160  
 Phe Glu Tyr Arg Asn Gly Ala Ala Val Thr Gly Val Thr Asp Arg Asn  
 165 170 175  
 Thr Val Val Asn His Phe Phe Asp Thr Pro Ile Arg Ala Arg Ser Ile  
 180 185 190  
 Ala Ile His Pro Leu Thr Trp Asn Gly His Ile Ser Leu Arg Cys Glu  
 195 200 205  
 Phe Tyr Thr Gln Pro Val Gln Ser Ser Val Thr Gln Val Gly Ala Asp  
 210 215 220  
 Ile Tyr Thr Gly Asp Asn Cys Ala Leu Asn Thr Gly Ser Gly Lys Arg  
 225 230 235 240  
 Glu Val Val Val Pro Val Lys Phe Gln Phe Glu Phe Ala Thr Leu Pro  
 245 250 255  
 Lys Val Ala Leu Asn Phe Asp Gln Ile Asp Cys Thr Asp Ala Thr Asn  
 260 265 270  
 Gln Thr Arg Ile Gly Val Gln Pro Arg Asn Ile Thr Thr Lys Gly Phe  
 275 280 285  
 Asp Cys Val Phe Tyr Thr Trp Asn Glu Asn Lys Val Tyr Ser Leu Arg  
 290 295 300  
 Ala Asp Tyr Ile Ala Thr Ala Leu Glu  
 305 310

## (2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 23 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TAYACTGAY AAYGTAAYTG GGT

23

## (2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single

80

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

MGDWSTATHG CATCAYCC

18

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Lys Val Glu Val Leu Pro Ala Leu Thr Asp Asn Tyr Met Tyr Leu  
1 5 10 15

Val Ile Asp Asp Glu Thr Lys Glu Ala Ala Ile Val Asp Pro Val Gln  
20 25 30

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Lys Ile Glu Leu Leu Pro Ala Leu Thr Asp Asn Tyr Met Tyr Leu  
1 5 10 15

Ile Ile Asp Glu Asp Thr Gln Xaa Ala Ala Val Val Asp Pro Val Gln  
20 25 30

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1161 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 158..937

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ACCGGCCCGG GTCATGGTGG TGGGCCGAGG GCTGCTCGGC CGCCGCAGCC TCGCCGCGCT 60  
GGGAGCCGCC TGGCCCCGCC GAGGCCTCGG TCCAGCCCTG CTGGAGTTT TCTGCCACAC 120  
AGATTTCGGG AAGAACCTGA CCGTGGACGA GGGCACC ATG AAG GTA GAG GTG CTG 175

81

Met Lys Val Glu Val Leu  
260

CCT GCC CTG ACC GAC AAC TAC ATG TAC CTG GTC ATT GAT GAT GAG ACC Pro Ala Leu Thr Asp Asn Tyr Met Tyr Leu Val Ile Asp Asp Glu Thr 265 270 275	223
AAG GAG GCT GCC ATT GTG GAT CCG GTG CAG CCC CAG AAG GTC GTG GAC Lys Glu Ala Ala Ile Val Asp Pro Val Gln Pro Gln Lys Val Val Asp 280 285 290 295	271
GCG GCG AGA AAG CAC GGG GTG AAA CTG ACC ACA GTG CTC ACC ACC CAC Ala Ala Arg Lys His Gly Val Lys Leu Thr Thr Val Leu Thr Thr His 300 305 310	319
CAC CAC TGG GAC CAT GCT GGC GGG AAT GAG AAA CTG GTC AAG CTG GAG His His Trp Asp His Ala Gly Gly Asn Glu Lys Leu Val Lys Leu Glu 315 320 325	367
TCG GGA CTG AAG GTG TAC GGG GGT GAC GAC CGT ATC GGG GCC CTG ACT Ser Gly Leu Lys Val Tyr Gly Gly Asp Asp Arg Ile Gly Ala Leu Thr 330 335 340	415
CAC AAG ATC ACT CAC CTG TCC ACA CTG CAG GTG GGG TCT CTG AAC GTC His Lys Ile Thr His Leu Ser Thr Leu Gln Val Gly Ser Leu Asn Val 345 350 355	463
AAG TGC CTG GCG ACC CCG TGC CAC ACT TCA GGA CAC ATT TGT TAC TTC Lys Cys Leu Ala Thr Pro Cys His Thr Ser Gly His Ile Cys Tyr Phe 360 365 370 375	511
GTG AGC AAG CCC GGA GGC TCG GAG CCC CCT GCC GTG TTC ACA GGT GAC Val Ser Lys Pro Gly Gly Ser Glu Pro Pro Ala Val Phe Thr Gly Asp 380 385 390	559
ACC TTG TTT GTG GCT GGC TGC GGG AAG TTC TAT GAA GGG ACT GCG GAT Thr Leu Phe Val Ala Gly Cys Gly Lys Phe Tyr Glu Gly Thr Ala Asp 395 400 405	607
GAG ATG TGT AAA GCT CTG CTG GAG GTC TTG GGC CGG CTC CCC CCG GAC Glu Met Cys Lys Ala Leu Leu Glu Val Leu Gly Arg Leu Pro Pro Asp 410 415 420	655
ACA AGA GTC TAC TGT GGC CAC GAG TAC ACC ATC AAC AAC CTC AAG TTT Thr Arg Val Tyr Cys Gly His Glu Tyr Thr Ile Asn Asn Leu Lys Phe 425 430 435	703
GCA CGC CAC GTG GAG CCC GGC AAT GCC GCC ATC CGG GAG AAG CTG GCC Ala Arg His Val Glu Pro Gly Asn Ala Ala Ile Arg Glu Lys Leu Ala 440 445 450 455	751
TGG GCC AAG GAG AAG TAC AGC ATC GGG GAG CCC ACA GTG CCA TCC ACC Trp Ala Lys Glu Lys Tyr Ser Ile Gly Glu Pro Thr Val Pro Ser Thr 460 465 470	799
CTG GCA GAG GAG TTT ACC TAC AAC CCC TTC ATG AGA GTG AGG GAG AAG Leu Ala Glu Glu Phe Thr Tyr Asn Pro Phe Met Arg Val Arg Glu Lys 475 480 485	847
ACG GTG CAG CAG CAC GCA GGT GAG ACG GAC CCG GTG ACC ACC ATG CGG Thr Val Gln Gln His Ala Gly Glu Thr Asp Pro Val Thr Thr Met Arg 490 495 500	895
GCC GTG CGC AGG GAG AAG GAC CAG TTC AAG ATG CCC CCG GAC Ala Val Arg Arg Glu Lys Asp Gln Phe Lys Met Pro Arg Asp 505 510 515	937
TGAGGCCGCC CTGCACCTTC AGCGGATTTG GGGATTAGGC TCGTTAGGT AACTGGCTTT	997
CCTGCTGGTC CGTGGCGGAA ATTCAGTCTT GATTTAACCT TAAATTTACA GCCCTTGGCT	1057
TGTGTTATCG GACGTTTTAA TGCATATTTA TAAGAGAAGT TTAACAAGTA TTTATCCCA	1117
TAAAAAGGGG GGGGCCGGTA CCAATTCCG CCTATAGTGA GTCG	1161

## (2) INFORMATION FOR SEQ ID NO:10:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 260 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```

Met Lys Val Glu Val Leu Pro Ala Leu Thr Asp Asn Tyr Met Tyr Leu
 1           5           10           15
Val Ile Asp Asp Glu Thr Lys Glu Ala Ala Ile Val Asp Pro Val Gln
 20           25           30
Pro Gln Lys Val Val Asp Ala Ala Arg Lys His Gly Val Lys Leu Thr
 35           40           45
Thr Val Leu Thr Thr His His His Trp Asp His Ala Gly Gly Asn Glu
 50           55           60
Lys Leu Val Lys Leu Glu Ser Gly Leu Lys Val Tyr Gly Gly Asp Asp
 65           70           75           80
Arg Ile Gly Ala Leu Thr His Lys Ile Thr His Leu Ser Thr Leu Gln
 85           90           95
Val Gly Ser Leu Asn Val Lys Cys Leu Ala Thr Pro Cys His Thr Ser
100           105           110
Gly His Ile Cys Tyr Phe Val Ser Lys Pro Gly Gly Ser Glu Pro Pro
115           120           125
Ala Val Phe Thr Gly Asp Thr Leu Phe Val Ala Gly Cys Gly Lys Phe
130           135           140
Tyr Glu Gly Thr Ala Asp Glu Met Cys Lys Ala Leu Leu Glu Val Leu
145           150           155           160
Gly Arg Leu Pro Pro Asp Thr Arg Val Tyr Cys Gly His Glu Tyr Thr
165           170           175
Ile Asn Asn Leu Lys Phe Ala Arg His Val Glu Pro Gly Asn Ala Ala
180           185           190
Ile Arg Glu Lys Leu Ala Trp Ala Lys Glu Lys Tyr Ser Ile Gly Glu
195           200           205
Pro Thr Val Pro Ser Thr Leu Ala Glu Glu Phe Thr Tyr Asn Pro Phe
210           215           220
Met Arg Val Arg Glu Lys Thr Val Gln Gln His Ala Gly Glu Thr Asp
225           230           235           240
Pro Val Thr Thr Met Arg Ala Val Arg Arg Glu Lys Asp Gln Phe Lys
245           250           255
Met Pro Arg Asp
260

```

## (2) INFORMATION FOR SEQ ID NO:11:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 271 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Pro Pro Asp Pro Ile Leu Gly Val Thr Glu Ala Phe Lys Arg Asp Thr  
 1 5 10 15  
 Asn Ser Lys Lys Met Asn Leu Gly Val Gly Ala Tyr Arg Asp Asp Asn  
 20 25 30  
 Gly Lys Ser Tyr Val Leu Asn Cys Val Arg Lys Ala Glu Ala Met Ile  
 35 40 45  
 Ala Ala Lys Lys Met Asp Lys Glu Tyr Leu Pro Ile Ala Gly Leu Ala  
 50 55 60  
 Asp Phe Thr Arg Ala Ser Ala Glu Leu Ala Leu Gly Glu Asn Ser Glu  
 65 70 75 80  
 Ala Phe Lys Ser Gly Arg Tyr Val Thr Val Gln Gly Ile Ser Gly Thr  
 85 90 95  
 Gly Ser Leu Arg Val Gly Ala Asn Phe Leu Gln Arg Phe Phe Lys Phe  
 100 105 110  
 Ser Arg Asp Val Tyr Leu Pro Lys Pro Ser Trp Gly Asn His Thr Pro  
 115 120 125  
 Ile Phe Arg Asp Ala Gly Leu Gln Leu Gln Ala Tyr Arg Tyr Tyr Asp  
 130 135 140  
 Pro Lys Thr Cys Ser Leu Asp Phe Thr Gly Ala Met Glu Asp Ile Ser  
 145 150 155 160  
 Lys Ile Pro Glu Lys Ser Ile Ile Leu Leu His Ala Cys Ala His Asn  
 165 170 175  
 Pro Thr Gly Val Asp Pro Arg Gln Glu Gln Trp Lys Glu Leu Ala Ser  
 180 185 190  
 Val Val Lys Lys Arg Asn Leu Leu Ala Tyr Phe Asp Met Ala Tyr Gln  
 195 200 205  
 Gly Phe Ala Ser Gly Asp Ile Asn Arg Asp Ala Trp Ala Leu Arg His  
 210 215 220  
 Phe Ile Glu Gln Gly Ile Asp Val Val Leu Ser Gln Ser Tyr Ala Lys  
 225 230 235 240  
 Asn Met Gly Leu Tyr Gly Glu Arg Ala Gly Ala Phe Thr Val Ile Cys  
 245 250 255  
 Arg Asp Ala Glu Glu Ala Lys Arg Val Glu Ser Gln Leu Lys Ile  
 260 265 270

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 629 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 56..595

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CATAATGAAA GTTATTTTCG TAGTTTTCG CATTGTATTA GTTACATTAT GGGCT ATG

84

Met

CCA TCA GAA GCT GGT AAA GAC CCA AAG ATT ACC AAT AAA GTA TTC TTT 106  
 Pro Ser Glu Ala Gly Lys Asp Pro Lys Ile Thr Asn Lys Val Phe Phe  
 265 270 275

GAT ATA GAA ATT GAT AAT AAA CCA GCA GGT AGA ATT GTA TTT GGT TTA 154  
 Asp Ile Glu Ile Asp Asn Lys Pro Ala Gly Arg Ile Val Phe Gly Leu  
 280 285 290

TAT GGA AAG ACA GTA CCA AAA ACA GTT GAA AAC TTT AGA GCA TTA TGT 202  
 Tyr Gly Lys Thr Val Pro Lys Thr Val Glu Asn Phe Arg Ala Leu Cys  
 295 300 305

ACT GGT GAA AAA GGT TTA GGT ACC AGT GGT AAA CCA TTA CAT TAT AAA 250  
 Thr Gly Glu Lys Gly Leu Gly Thr Ser Gly Lys Pro Leu His Tyr Lys  
 310 315 320 325

GAT AGT AAA TTC CAT CGT ATC ATT CCA AAC TTT ATG ATT CAA GGT GGT 298  
 Asp Ser Lys Phe His Arg Ile Ile Pro Asn Phe Met Ile Gln Gly Gly  
 330 335 340

GAT TTC ACA AGA GGT GAT GGT ACT GGT GGT GAA TCA ATT TAT GGT AAA 346  
 Asp Phe Thr Arg Gly Asp Gly Thr Gly Gly Glu Ser Ile Tyr Gly Lys  
 345 350 355

AAA TTC AAT GAT GAA AAC TTC AAA ATT AAA CAC TCC AAA CCA GGT CTT 394  
 Lys Phe Asn Asp Glu Asn Phe Lys Ile Lys His Ser Lys Pro Gly Leu  
 360 365 370

TTA TCA ATG GCT AAC GCT GGT CCA AAC ACT AAT GGT TCA CAA TTC TTT 442  
 Leu Ser Met Ala Asn Ala Gly Pro Asn Thr Asn Gly Ser Gln Phe Phe  
 375 380 385

ATT ACT ACC GTT GTT ACT TCA TGG TTA GAT GGT CGT CAT ACT GTT TTT 490  
 Ile Thr Thr Val Val Thr Ser Trp Leu Asp Gly Arg His Thr Val Phe  
 390 395 400 405

GGT GAA GTT ATT GAA GGT ATG GAT ATT GTT AAA CTC CTT GAA TCC ATT 538  
 Gly Glu Val Ile Glu Gly Met Asp Ile Val Lys Leu Leu Glu Ser Ile  
 410 415 420

GGT TCC CAA TCT GGA ACA CCA AGT AAA ATT GCT AAA ATC TCA AAC TCT 586  
 Gly Ser Gln Ser Gly Thr Pro Ser Lys Ile Ala Lys Ile Ser Asn Ser  
 425 430 435

GGT GAA TTA TAAATAAAAT AAAACCAAC CAAATAAAAT AAAT 629  
 Gly Glu Leu  
 440

## (2) INFORMATION FOR SEQ ID NO:13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 180 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Met Pro Ser Glu Ala Gly Lys Asp Pro Lys Ile Thr Asn Lys Val Phe  
 1 5 10 15

Phe Asp Ile Glu Ile Asp Asn Lys Pro Ala Gly Arg Ile Val Phe Gly  
 20 25 30

Leu Tyr Gly Lys Thr Val Pro Lys Thr Val Glu Asn Phe Arg Ala Leu  
 35 40 45

Cys Thr Gly Glu Lys Gly Leu Gly Thr Ser Gly Lys Pro Leu His Tyr  
 50 55 60

85

Lys Asp Ser Lys Phe His Arg Ile Ile Pro Asn Phe Met Ile Gln Gly  
 65 70 75 80  
 Gly Asp Phe Thr Arg Gly Asp Gly Thr Gly Gly Glu Ser Ile Tyr Gly  
 85 90 95  
 Lys Lys Phe Asn Asp Glu Asn Phe Lys Ile Lys His Ser Lys Pro Gly  
 100 105 110  
 Leu Leu Ser Met Ala Asn Ala Gly Pro Asn Thr Asn Gly Ser Gln Phe  
 115 120 125  
 Phe Ile Thr Thr Val Val Thr Ser Trp Leu Asp Gly Arg His Thr Val  
 130 135 140  
 Phe Gly Glu Val Ile Glu Gly Met Asp Ile Val Lys Leu Leu Glu Ser  
 145 150 155 160  
 Ile Gly Ser Gln Ser Gly Thr Pro Ser Lys Ile Ala Lys Ile Ser Asn  
 165 170 175  
 Ser Gly Glu Leu  
 180

## (2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Asn Ser Ile Leu Asn Phe Ser Asn Ser Lys  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

His Phe Val Xaa Ile Ser Thr Gln Gly Arg Gly Asp His Asp Gln Xaa  
 1 5 10 15

Val Thr Xaa Tyr  
 20

## (2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear



(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Gly Thr Gly Ser Arg Thr Ile Val  
1 5

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 16 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Asp Ala Ser Arg Phe Asp Gly Ser Trp Ser Ser Xaa Val Leu Asp Lys  
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Leu Arg Tyr Thr Leu Asp Asn Val Asn Trp Val Glu Tyr Asn Asn Gly  
1 5 10 15

Glu Ile Asn Ala Asn Lys  
20

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 17 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Xaa Arg Ser Ile Ala Ile His Pro Thr Tyr Asn Asn His Ile Ser Ile  
1 5 10 15

Arg

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 14 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

87

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Asp Asn Gly Gln Met Arg Trp Glu Gly Lys Ser Glu Asn Ile  
 1 5 10

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Asp Leu Thr Phe Ile Thr Trp Gly Asn Asn Ala Val Tyr  
 1 5 10

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 48 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Asp Ser Val Lys His Phe Val Ala Ile Ser Thr Gln Gly Arg Gly Asp  
 1 5 10 15

His Asp Gln Trp Val Thr Ser Tyr Lys Leu Arg Tyr Thr Leu Asp Asn  
 20 25 30

Val Asn Trp Val Glu Tyr Asn Asn Gly Glu Ile Ile Asn Ala Asn Lys  
 35 40 45

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 35 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GACTCGAGTC GACATCGATT TTJTTTTTTT TTTT

35

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 18 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GACTCGAGTC GACATCGA

18

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 32 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Met	Lys	Val	Glu	Val	Leu	Pro	Ala	Leu	Thr	Asp	Asn	Tyr	Met	Tyr	Leu
1				5					10					15	
Val	Ile	Asp	Asp	Glu	Thr	Lys	Glu	Ala	Ala	Ile	Val	Asp	Pro	Val	Gln
		20					25						30		

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 21 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Tyr	Xaa	Ile	Gly	Glu	Pro	Thr	Val	Pro	Ser	Thr	Leu	Ala	Glu	Glu	Phe
1				5					10					15	
Thr	Tyr	Asn	Pro	Phe											
			20												

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 21 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GCTTACNGAY AAYTAYATGT A

21

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 20 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GAYGAYGARA CNAARGARGC

20

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 16 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

ATHGGGARCC ACGTGG

16

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 16 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TADCCCTYGG TGCAGG

16

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 37 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GGCCACGCGT CGACTAGTAC TTTTTTTTTT TTTTTT

37

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 32 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CUACUACUAC UAGGCCACGC GTCGACTAGT AC

32

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 14 amino acids

90

(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Gly Lys Asp Pro Lys Ile Thr Asn Lys Val Phe Phe Asp Glu  
1                   5                   10

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

His Gly Val Lys  
1

What is claimed is:

1. A substantially pure bisphosphonate binding protein or a fragment or homologue or mutant thereof.
2. The protein of Claim 1 which mediates the physiological effects of bisphosphonates in vivo.
3. The protein of Claim 1 which binds any or all of pyridoxal phosphate, O-phosphorylethanolamine, O-phosphorylcholine, phosphatidyl ethanolamine or phospholipid bisphosphonate analogues.
4. The protein of Claim 1 prepared by recombinant methods, which is conjugated to an antibody, to a fusion protein, to a mimetic or mimetic analogue or to a solid support..
5. The protein Claim 1 chosen from DPI (SEQ.ID. NO.2); hDPI (SEQ.ID. NO.10); the Dictyostelium homologue thereof (Dd-hDPI); DdCyP2 (SEQ.ID. NO.13); or homologues thereof; fragments thereof; and proteins containing any of the preceding which substantially retain bisphosphonate binding activity.
6. A method for purifying a bisphosphonate binding protein of claim 1, comprising the steps of:
  - (a) linking bisphosphonate to a chromatography column to produce an affinity column;
  - (b) loading material containing impure bisphosphonate binding material onto the affinity column such that it becomes bound thereto; and
  - 5 (c) selectively eluting the binding protein from the affinity column in a more purified form.
7. A method for producing a bisphosphonate binding protein of Claim 1, comprising the steps of:
  - (a) providing a Dictyostelium mutant which expresses a mutated bisphosphonate-binding-protein gene;
  - (b) cloning the wild-type gene corresponding to that mutated in step (a) to produce a cloned bisphosphonate-binding-protein gene; and
  - (c) expressing the cloned bisphosphonate-binding-protein gene to produce the bisphosphonate binding protein.

8. The protein of Claim 1 comprised in a pharmaceutical excipient.
  9. Isolated DNA, vector or host cell which encodes the proteins of Claim 1 comprising the groups consisting of SEQ.ID NO. 1; SEQ.ID NO. 9 and SEQ.ID NO. 12.
  10. A method for evaluating the therapeutic activity of a bisphosphonate comprising the steps of:
    - (a) contacting the bisphosphonate with the bisphosphonate binding protein of Claim 1; and
    - (b) measuring the binding affinity of the bisphosphonate binding protein for the bisphosphonate.
- 5
11. An antibody which binds to the binding protein of Claim 1, or a derivative thereof.
  12. A test kit comprising (i) the bisphosphonate binding protein of Claim 1 bound to a solid support.
  13. A method of diagnosing calcium metabolism disorders in a mammal using a bisphosphonate binding protein of claim 1, an antibody of the bisphosphonate binding protein, or an antagonist of a bisphosphonate binding protein.
  14. A method of treating a calcium metabolism disorder using a bisphosphonate binding protein of Claim 1, antibody thereto, or antagonist wherein the treatment for the regulation of bone metabolism, hypercalcaemia, bone metastases and osteoporosis.

15. A method of treating a calcium metabolism disorder using a bisphosphonate binding protein of Claim 1, antibody thereto, or antagonist wherein the therapy involves the regulation of bone metabolism via interaction with cyclosporin.



# INTERNATIONAL SEARCH REPORT

Intern: al Application No  
PCT/US 98/02709

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/12 C12N15/55 C12N15/80 C12N15/85 C12N9/16  
C12N1/15 C07K14/37 C07K14/47 C07K1/22 C07K16/14  
C07K16/18 G01N33/50 A61K38/16 A61K38/17

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K G01N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	FUKUZAWA M. & OCHIAI H.: "Monoclonal antibodies against Discoidin I and Discoidin II of the cellular slime mold D. discoideum." J. BIOCHEM., vol. 103, 1988, pages 884-888, XP002068194 see abstract see figures 1-3 --- -/-	1-3,5,9, 11

<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C.	<input type="checkbox"/> Patent family members are listed in annex.
<p>* Special categories of cited documents :</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"Z" document member of the same patent family</p>	
Date of the actual completion of the international search	Date of mailing of the international search report
18 June 1998	06.07.98
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer  Galli, I

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/02709

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	FUKUZAWA M. & OCHIAI H.: "Molecular cloning and characterization of the cDNA for Discoidin II of D. discoideum" PLANT CELL PHYSIOL., vol. 37, no. 4, June 1996, pages 505-514, XP002068547 cited in the application see abstract see figures 2,3 ---	1-3,5,9, 11
X	RIDDERSTRÖM M. ET AL.: "Molecular cloning, heterologous expression and chracterization of human glyoxalase II" J. BIOL. CHEM., vol. 271, no. 1, 5 January 1996, pages 319-323, XP002068197 see Fig. 1: hDPI identical to human glyoxalase II ---	1-3,5,9
X	DALEY-YATES P.T. ET AL.: "Plasma protein binding of APD: role of calcium and transferrin" CHEM. BIOL. INTERACT., vol. 81, no. 1-2, January 1992, pages 79-89, XP002068193 see abstract ---	1-3
A	ROGERS M.J. ET AL.: "Structure-activity relationship of new heterocycle-containing bisphosphonates as inhibitors of bone resorption and as inhibitors of D. discoideum amoebae" MOL. PHARM., vol. 47, no. 2, February 1995, pages 398-402, XP002068195 see abstract ---	1-15
A	VITTÉ C. ET AL.: "Bisphosphonates induce osteoblasts to secrete an inhibitor of osteoclast-mediated resorption" ENDOCRINOLOGY, vol. 137, no. 6, June 1996, pages 2324-2333, XP002068196 see abstract ---	1-15
A	HALL T.J. ET AL.: "The role of reactive oxygen intermediates in osteoclastic bone resorption." BIOCHEM. BIOPHYS. RES. COMM., vol. 207, no. 1, 6 February 1995, pages 280-287, XP002068360 see abstract: link between free radicals, antioxidant systems and osteoporosis -----	5,9

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US 98/02709

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
Although claims 14 and 15 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.